

**Univerzita Karlova  
Přírodovědecká fakulta**

Studijní program: Demografie

Studijní obor: Demografie se sociální geografii



**Adéla Kadlecová**

Aplikace víceúrovňového modelování v geografii: příklad nezaměstnanosti  
Application of Multilevel Models in Geography: Example of Unemployment

Bakalářská práce

Vedoucí práce: RNDr. Pavlína Netrdová, Ph.D.

Praha, 2018

## Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 31. 7. 2018

.....

Adéla Kadlecová

Na tomto místě bych chtěla poděkovat především RNDr. Pavlíně Netrdové, Ph.D. za vedení této bakalářské práce, kterému v její prospěch věnovala nespočet hodin práce, mnoho cenných rad a neutuchající nadšení. Za velmi vřelou ochotu bych ráda poděkovala také Ing., Mgr. Martinu Zelenému, Ph.D.

## **Aplikace víceúrovňového modelování v geografii: příklad nezaměstnanosti**

### **Abstrakt**

Víceúrovňové modely představují efektivní metodologický nástroj pro analýzu hierarchicky uspořádaných dat. Jedná se o kvantitativní metodu založenou na regresní analýze, jejíž užívání není v současné době v geografii příliš rozšířené. Vzhledem k této skutečnosti si tak tato práce klade za cíl zejména podrobné představení statistické metody víceúrovňového modelování, zahrnující jak seznámení se všemi jejími možnostmi a omezeními, tak také uvedení příkladu aplikace této metody v konkrétním geografickém výzkumu z oblasti nezaměstnanosti. Smyslem této analýzy je identifikace jednotlivých individuálních a regionálních proměnných souvisejících s nezaměstnaností a odhalení efektu těchto proměnných na nezaměstnanost.

**Klíčová slova:** víceúrovňové modelování, kvantitativní metody, nezaměstnanost, Česko

## **Application of Multilevel Modeling in Geography: Example of Unemployment**

### **Abstract**

Multilevel models are an effective methodological tool for analyzing hierarchically structured data. It is a quantitative method based on regression analysis whose use in geography is currently not so widespread. In view of this fact this work is aimed especially at a detailed introduction of the statistical method of multilevel modeling, including both an introduction to all its possibilities and limitations, as well as an example of application of this method in the specific geographic research from the area of unemployment. The purpose of this analysis is to identify particular individual and regional variables related to unemployment and to uncover the effect of these variables on unemployment.

**Keywords:** multilevel modeling, quantitative methods, unemployment, Czechia

## OBSAH

<b>PŘEHLED POUŽITÝCH ZKRATEK.....</b>	<b>6</b>
<b>SEZNAM OBRÁZKŮ.....</b>	<b>7</b>
<b>SEZNAM TABULEK .....</b>	<b>8</b>
<b>1 ÚVOD .....</b>	<b>9</b>
<b>2 KVANTITATIVNÍ METODY V SOCIÁLNÍ GEOGRAFII.....</b>	<b>11</b>
<b>3 TEORETICKÉ ASPEKTY VÍCEÚROVŇOVÉHO MODELOVÁNÍ .....</b>	<b>15</b>
3.1 TEORETICKÁ MOTIVACE APLIKACE VÍCEÚROVŇOVÝCH MODELŮ .....	16
3.2 STATISTICKÁ MOTIVACE APLIKACE VÍCEÚROVŇOVÝCH MODELŮ .....	18
3.3 MOŽNOSTI A OMEZENÍ VÍCEÚROVŇOVÝCH MODELŮ .....	20
3.4 KONSTRUKCE VÍCEÚROVŇOVÝCH MODELŮ .....	24
<b>4 PŘÍKLADY APLIKACÍ VÍCEÚROVŇOVÝCH MODELŮ .....</b>	<b>29</b>
<b>5 VÍCEÚROVŇOVÉ MODEL Y A NEZAMĚSTNANOST.....</b>	<b>33</b>
5.1 PŘÍPRAVA DAT PRO ANALÝZU .....	35
5.2 ANALÝZA RIZIKA NEZAMĚSTNANOSTI NA INDIVIDUÁLNÍ ÚROVNI .....	40
5.3 ODHALENÍ REGIONÁLNÍCH VlivŮ NA RIZIKO NEZAMĚSTNANOSTI .....	45
5.4 VÍCEÚROVŇOVÁ ANALÝZA RIZIKA NEZAMĚSTNANOSTI .....	48
<b>6 ZÁVĚR .....</b>	<b>50</b>
<b>SEZNAM LITERATURY .....</b>	<b>53</b>
<b>PŘÍLOHY .....</b>	<b>61</b>

## PŘEHLED POUŽITÝCH ZKRATEK

CIRCABS	<i>Communication and Information Resource Centre for Administrations, Businesses and Citizens</i>
ČSÚ	Český statistický úřad
EU-SILC	Výběrové šetření Životní podmínky ( <i>European Union – Statistics on Income and Living Conditions</i> )
ICC	Vnitrotřídní koeficient korelace ( <i>intra-class correlation coefficient</i> )
RSO	
SLDB 2011	Registr sčítacích obvodů a budov Sčítání lidu, domů a bytů 2011

## SEZNAM OBRÁZKŮ

<b>Obrázek 1:</b> Regresní přímky s odlišnými konstantami .....	28
<b>Obrázek 2:</b> Regresní přímky s odlišnými konstantami i směrnici .....	28
<b>Obrázek 3:</b> Model rizika nezaměstnanosti .....	34
<b>Obrázek 4:</b> Histogram znázorňující rozložení hodnot ICC; základní vícestupňové výběrové soubory; 100 souborů .....	47
<b>Obrázek 5:</b> Graf znázorňující hodnoty ICC; vícestupňové výběrové soubory „kraj“; 10 souborů .....	47
<b>Obrázek 6:</b> Graf znázorňující hodnoty ICC; vícestupňové výběrové soubory „obec“; 10 souborů .....	48

## SEZNAM TABULEK

<b>Tabulka 1:</b> Individuální determinanty nezaměstnanosti analyzované v jednotlivých studiích .....	36
<b>Tabulka 2:</b> Seznam analyzovaných proměnných .....	41
<b>Tabulka 3:</b> Kontingenční koeficienty (Pearson Chi-Square) vyjadřující závislost mezi vybranými proměnnými .....	42
<b>Tabulka 4:</b> Výstupy frekvenčních a párových analýz rizika nezaměstnanosti na individuální úrovni; základní soubor; 2 232 757 osob .....	43
<b>Tabulka 5:</b> Model binární logistické regrese; základní soubor; 2 232 757 osob .....	44
<b>Tabulka 6:</b> Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující krajskou příslušnost; základní soubor; 2 232 757 osob .....	45
<b>Tabulka 7:</b> Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující velikostní kategorie obcí; základní soubor; 2 232 757 osob .....	46



## Kapitola 1

### Úvod

Zabývat se v rámci vysokoškolské kvalifikační práce či případně dále v samotném vědeckém výzkumu problematikou zejména metodologickou není na poli geografie příliš běžnou praxí. Přesto, nebo spíše právě proto, jsem se rozhodla věnovat svoji bakalářskou práci bližšímu představení jedné konkrétní statistické metody, jejíž znalost a následné vhodné použití může být nejen pro geografický výzkum značným přínosem, a to metody víceúrovňového modelování.

Víceúrovňové modely jsou postupně rozvíjeny již několik desetiletí, avšak k tomuto rozvoji dochází převážně v zahraničí. Jejich rozsáhlejší aplikaci a praktickému využití se v českém prostředí zatím nedostalo, proto můžeme o používání této metody u nás hovořit vesměs jako o marginálním (Soukup 2006). Přestože je zde tento stav reflektující zejména pomalý rozvoj a integraci nových metod do aktuálních výzkumných prací v našem prostředí demonstrován pouze na úzce vymezeném případě jedné konkrétní metody, je bohužel pravdou, že reálně dokonale ilustruje stav, ve kterém se nachází celá současná vědecká a statistická geografická metodologie. Té je mezi odbornou veřejností, a to podle mého názoru neprávem, věnována stále poměrně malá pozornost. Z tohoto důvodu pak často nedochází k jejímu adekvátnímu rozvoji, který je v porovnání s vývojem empiricky zaměřených geografických prací a výzkumů nesrovnatelný. A to paradoxně i přesto, že veškeré tyto konkrétní tematické studie by se bez vhodného metodologického aparátu jen stěží obešly a jejichž výstupy, které jsou předmětem hodnocení jejich kvality, se často přímo odvíjí právě od použitých metod, a tudíž i významně souvisí se stupněm jejich poznání. Mnohými tematicky zaměřenými geografiy je však určitý stupeň metodického poznání vnímán jako samozřejmost, jako nástroj, který je jejich studiu neustále k dispozici.

Jedním z příkladů nesamozřejmosti tohoto stavu zdánlivé běžné všudypřítomnosti vhodných metod by snad mohla být tato práce, jejímž cílem je podrobné představení v geografickém výzkumu zatím spíše okrajově používané statistické metody víceúrovňového modelování.

Metoda víceúrovňového modelování bude představena především skrze vyzdvižení jejích přínosů a upozornění na úskalí, které s sebou může její použití v geografickém výzkumu přinášet. Řada těchto úskalí vychází z faktu, že víceúrovňové modelování je metodou kvantitativní, a tudíž se při jejím využití potýkáme se všemi tradičními problémy spojenými s aplikací kvantitativních metod v geografii, potažmo v sociálních vědách obecně. Proto je hned následující kapitola zasvěcena právě tématu využití kvantitativních metod v sociální geografii. Svým zaměřením se však jedná spíše o kapitolu úvodní, která se snaží postihnout tuto problematiku opravdu velice stručně. Jejím účelem totiž není komplexní, podrobné a vyčerpávající pojetí celé této problematiky, nýbrž spíše ukotvení samotné metody víceúrovňového modelování v rámci širší typologie metod používaných v nejrůznějších geografických výzkumech a představení nejdůležitějších, zejména praktických aplikačních a interpretačních důsledků tohoto zařazení. Po této kapitole již následuje část věnující se konkrétně víceúrovňovým modelům zahrnující jak kapitoly výhradně teoretické, popisující podstatu této metody, jednotlivé motivace k jejímu použití či samotnou konstrukci modelů, tak také kapitoly zaměřené spíše na seznámení s některými již uskutečněnými studiemi aplikujícími víceúrovňově konstruované modely jak v oblasti geografie, tak také na poli jiných vědních oborů.

Speciální pozornost je poté věnována konkrétnímu využití metody víceúrovňového modelování při analýze rizika nezaměstnanosti. Přestože problematika nezaměstnanosti není příliš typickou ilustrační oblastí, ve které by byly víceúrovňově konstruované modely hojně využívány, umožnila jejich aplikaci právě v této sféře dostupnost datového souboru s individuálními daty o jednotlivých osobách včetně jejich regionální příslušnosti. Kromě metodologické roviny zahrnující zejména samotnou ukázkou možností využití metody víceúrovňového modelování v geografickém výzkumu, lze přínos aplikačního příkladu identifikovat také v empirické rovině. Tou je zejména identifikace jednotlivých individuálních a regionálních proměnných souvisejících s nezaměstnaností a odhalení velikosti efektu těchto proměnných na nezaměstnanost. V závěrečné kapitole jsou pak výsledky těchto empirických analýz shrnuty a kriticky zhodnoceny.

## Kapitola 2

### Kvantitativní metody v sociální geografii

Sociální geografii bývá někdy obtížné přesně identifikovat a ukotvit v systému věd. Možností, jak jednotlivé vědní disciplíny vůbec klasifikovat, je mnoho. Jednou z nich je na první pohled jednoduché rozdělení na vědy exaktní a společenské, které vychází z díla Thomase Kuhna (Disman 2002). Toto dělení, podle kterého sociální geografie náleží k vědám společenským, je v tomto případě na místě, jelikož skrze charakter konkrétních závěrů velmi dobře reflektuje možnosti kvantitativního přístupu v obou skupinách věd. Charakter společenskovědních závěrů podle Dismana (2002) určují zejména tři jeho základní vlastnosti, a to že závěr je vždy stochastického, pravděpodobnostního rázu, dále že neplatí univerzálně, nýbrž pouze v prostředí, ze kterého pocházela data, a že je také velice obtížné, často dokonce nemožné prokázat jeho kauzální charakter.

Problematika určité proměnlivosti společenskovědních závěrů úzce souvisí s konceptem, který Ashby (1965, cit. v Disman 2002) nazývá přirozený systém a označuje jím soubor proměnných, které jsou navzájem propojeny mnoha vztahy. Právě tyto přirozené systémy jsou zpravidla objektem zájmu vědeckých prací, jsou tím, co zkoumáme, a proto výzkumné závěry ve velké míře závisejí na charakteru těchto systémů a to zejména na jejich velikosti a složitosti. Obecně můžeme říci, že přirozené systémy v sociálních vědách jsou mnohem rozsáhlejší než systémy ve vědách exaktních, což značně zesložituje proces společenskovědního bádání. V jeho průběhu se neobejdeme bez modelů popisujících skutečné okolní prostředí, které však tuto realitu značně zjednodušují. Disman (2002) identifikuje čtyři takovéto úrovně, na kterých dochází k nezbytné redukci informace. První dvě úrovně, kterými jsou redukce počtu pozorovaných proměnných a redukce počtu analyzovaných vztahů mezi nimi, přímo vyplývají z velikosti a složitosti přirozeného systému a jsou jasným dokladem skutečnosti, „že v sociálních vědách prakticky nikdy

nebudeme pracovat s úplně popsaným přirozeným systémem“ (Disman 2002, s. 25). Dvěma zbývajícími úrovněmi jsou poté redukce populace na vzorek a redukce časového kontinua na časový bod.

Problematika týkající se pojetí času souvisí do určité míry i s druhou výše zmiňovanou charakteristikou společenskovedních závěrů a to s jejich neuniverzální platností. Tyto závěry jsou totiž jak právě v časovém, tak v prostorovém pojetí platné pouze v prostředí, ze kterého pocházela data, což v důsledku značí, že v rámci společenskovedních výzkumů není v podstatě možné úplně abstrahovat od časoprostorových souřadnic (Petrusek 1993). Těžko můžeme například model vysvětlující příčiny nezaměstnanosti obyvatel v socialistickém Československu aplikovat do kapitalistického světa či do současné doby.

Třetí nedokonalostí a překážkou konstrukce naprosto přesného modelu je velice obtížné, často dokonce nemožné prokázání kauzálního charakteru společenskovedních závěrů. Složitost doložení takovéto příčinné závislosti mezi zkoumanými jevy je v podstatě přímo úměrná velikosti a složitosti přirozeného systému těchto jevů, o němž tedy již víme, že je v sociálních vědách velice rozsáhlý. Ve většině případů tak nemůžeme hovořit o jedné příčině zkoumaného jevu, nýbrž o komplexu příčin, při jehož synergickém působení se závisle proměnná nemění stejně v každém jednotlivém případě, ale v průměru, což neodpovídá závislosti kauzální, ale závislosti takzvané volné, statistické (Řeháková 1982).

Kromě těchto omezení vztahujících se k sociálním vědám obecně limituje kvantitativní metody v sociální geografii paradoxně také to, co dělá geografii geografii. Povaha geografických systémů, které jsou objektem studia geografie, je podle Hamplovy klasifikace reálných systémů (Hampl 1998) velice komplexní, čili velikost a zejména složitost přirozeného systému v geografii je obrovská. V neposlední řadě je to také prostorový charakter většiny geografických dat. Jeho výhodou je zajisté existence informace udávající polohu zkoumaného jevu, která je nadstavbou k standardní informaci atributové vypovídající o jeho charakteristikách, na stranu druhou však s sebou prostorovost může přinášet určité komplikace. Ty poté kromě užití jiných než tradičních statistických metod vyžadují také obezřetnější interpretaci získaných výsledků (Spurná 2006). Hlavními příčinami této obezřetnosti jsou zejména prostorová nestacionarita, neboli nestálost vztahu mezi proměnnými v rámci prostoru, prostorová závislost geografických dat a jejich agregace a s ní spojený problém ekologické chyby,

který budeme podrobněji diskutovat v následující kapitole v souvislosti s teoretickou motivací použití metod víceúrovňového modelování.

Obecně řečeno to však jsou právě prostorové aspekty geografických dat, které stále silněji přitahují pozornost mnoha geografů a které jsou v jistém smyslu příčinou a spouštěčem jakési rehabilitace kvantitativní geografie, ke které v poslední době dochází skrze vznik a následnou aplikaci nejrozličnějších nových prostorových metod (Fotheringham, Brunson, Charlton 2000). Celý tento přístup, který je založen na využití prostorového aspektu dat, bývá často souhrnně nazýván prostorová analýza dat a je úzce spjat s v současnosti populárním konceptem takzvané moderní kvantitativní geografie. Jedním z hlavních cílů tohoto konceptu je vznik nových metod vyvinutých speciálně pro geografické výzkumy právě s ohledem na ony prostorové aspekty, které postupně nahrazují dříve hojně používané metody odvozované z jiných disciplín a následně modifikované pro data nesoucí prostorovou informaci (Spurná 2006).

Typickým příkladem takovýchto metod je skupina geograficky vážených modelů (*geographically weighted models*), jejichž společným jmenovatelem je využití lokálního vážení jako součásti odhadu modelu. Tento přístup byl navržen na konci 70. let 20. století profesorem Clevelandem (1979, cit. v Timofeeva a Tesselkina 2017, s. 664) a dnes je základem celé řady jak tradičních, tak také neustále nově vznikajících, široce používaných metod prostorové analýzy dat. Mezi nejtradičnější a také nejčastěji používané patří metoda geograficky vážené regrese (*geographically weighted regression*) (Fotheringham, Brunson, Charlton 2002). Ta je založena, jak i samotný název napovídá, na tradičním regresním rámci, který je však rozšířen o pohyblivé lokální regresní parametry. Samotný výpočet lokálních regresních koeficientů je realizován právě skrze geografické vážení, které vychází ze vzdálenosti datového bodu od regresního bodu, přičemž se snižující se vzdáleností se daná váha zvyšuje (Spurná 2006). Výběr konkrétního typu vážící funkce označují spolu s určením šířky pásma Fotheringham, Brunson a Charlton (2002) jako klíčové body použití této metody, přičemž zejména volba šířky pásma má na výsledky získané metodou geograficky vážené regrese zásadní vliv. Kromě této tradiční metody se můžeme setkat také s použitím dalších geograficky vážených modelů, které jsou ve své podstatě modifikací geograficky vážené regrese a mezi které patří například geograficky vážená logistická regrese (*geographically weighted logistic regression*) či geograficky

vážená faktorová analýza (*geographically weighted factor analysis*) (Fotheringham, Brunson, Charlton 2002).

Významným faktorem rozvoje prostorové analýzy dat je bezpochyby také progresivní vývoj moderních informačních technologií. Díky tomuto vývoji je v geografickém výzkumu možné nejenom masivně využívat takzvaná „velká data“, ale také analyzovat vstupní informace pomocí zcela nových metod. Jednou z takovýchto novodobých metod je metoda nazývaná kódový shluk (*code cluster*), kterou ve svém výzkumu využila dvojice původem jihokorejských autorů Hong a Jung (2017). Díky této metodě, která v sobě kombinuje jak přístup kvantitativní, tak i kvalitativní, čímž reflektuje další dnes oblíbený metodologický trend, je kromě analýzy prostorových charakteristik a rozložení jednotlivých clusterů možné vysvětlit také geografické rozdíly v kvalitativních znacích konkrétních clusterů (Hong, Jung 2017). Pro jejich vymezení byla v této konkrétní studii použita data z webové služby Foursquare spadající do kategorie služeb často označovaných jako LSBN (*location-based social network*), která současně s údaji o poloze využívá také prvky sdílení známé ze sociálních sítí. Každý měsíc ji využije na padesát milionů uživatelů, kteří, nejčastěji pomocí aplikace ve svém chytrém telefonu, označí restaurace, kavárny, obchody, parky či jiná veřejná místa, která navštívili, mohou je ohodnotit a přidat fotografii a díky Foursquare sdílí tyto návštěvy se svými přáteli (Lutonský 2011). Na základě tohoto uživateli vytvořeného obsahu pak byly v rámci tohoto výzkumu provedeny ony kombinované analýzy, které hodnotily jak kvantitativní prostorové rozložení kaváren v metropolitní oblasti Seattlu, tak také jejich kvalitativní charakteristiky. Prostorové rozložení jednotlivých clusterů poté zobrazili pomocí teplotní mapy (*heatmap*), která je široce využívaným prostředkem geovizualizace právě při analýzách sociálních dat s prostorovou informací.

Přestože by se podle předcházejících odstavců mohlo zdát, že prostorovost naprosto pohltila současnou kvantitativní geografii, není tomu tak zcela, neboť například právě metoda, která je alfou a omegou celé této práce a která si i v rámci geografických výzkumů pomalu vydobývá své oprávněné místo na slunci, je metodou neprostorovou. Její bližší představení tak bude hlavní náplní následujících kapitol.

## Kapitola 3

### Teoretické aspekty víceúrovňového modelování

Termín víceúrovňové modelování či jeho anglický ekvivalent *multilevel modeling* nepatří mezi pojmy, které by byly v současné době příslušníkům české odborné geografické veřejnosti příliš známé. Přestože jsou metody skrývající se za těmito označeními zobecněním standardních metod regresní analýzy, jejich věhlasu zdaleka nedosahují. Tuto skutečnost nechává ze svého veskrze úvodního a průkopnického článku týkající se této problematiky v českém prostředí tušit Soukup (2006), podle něhož bylo u nás užívání metod víceúrovňového modelování v té době zatím takřkajíc „v plenkách“. Bohužel ani Voda (2013) si o celých sedm let později v úvodu svého příspěvku nemohl odpustit řádky, ve kterých se směsí smutku, drobné vyčítavosti a zoufalství poznamenává, že ač se víceúrovňové modelování řadí v západních výzkumech ke standardně využívaným metodám v rámci většiny vědeckých disciplín a v některých oborech dokonce dominuje, v žádné geografické studii vztažené k území Česka nebylo doposud použito. V zahraničí se do širšího povědomí dostaly tyto metody již ve druhé polovině 80. let 20. století a zejména pak od poloviny 90. let jsou stále masivněji využívány právě v nejrůznějších společenskovědních výzkumech. K tomuto rozšíření významně přispěla také implementace víceúrovňového modelování do komerčně nabízených a široce využívaných statistických softwarů, jakými jsou například SAS, STATA či SPSS (Soukup 2006). Vedle této snazší dostupnosti jsou to hlavně nezpochybnitelné přínosy a nové možnosti metod víceúrovňového modelování, které měly klíčový význam pro získání a následné ukotvení pevné pozice víceúrovňových modelů na poli vědeckých geografických metod. Mnozí autoři (Bock 1989; Kreft, de Leeuw 1998; Luke 2004; Soukup 2006; Heck a kol. 2012) hovoří o nejrůznějších motivacích, které zahrnují tyto jednotlivé přínosy a které jsou právě klíčem k onomu běžnějšímu používání metod víceúrovňového modelování. V praxi se

můžeme setkat s různými členěními těchto motivací, přičemž většina z nich je ve své podstatě obdobou nejjednoduššího dělení do dvou kategorií, které například Luke (2004, cit. V Pecha 2009, s. 5) nazývá teoretická a čistě statistická motivace.

### 3.1 Teoretická motivace aplikace víceúrovňových modelů

Základním termínem, který je spojen s teoretickou motivací aplikace víceúrovňových modelů a který ji stručně a výstižně charakterizuje, je kontextuální analýza. Tento pojem pocházející původně ze sociologie označuje typ analýzy, který je založen na hledání širších souvislostí zkoumaných jevů (Buriánek 2017). K sociologii a k sociálním vědám obecně tyto širší souvislosti a kontextovost neodmyslitelně patří, jelikož jednou z podstatných vlastností jejich přirozeného systému je jeho velikost a velká složitost, kterou jsme podrobněji diskutovali v předcházející kapitole. Jejím klíčovým důsledkem pro analýzu dat pocházejících ze společenskovedního prostředí je nutnost uvědomění si, že tato data často mají určitou hierarchickou strukturu. Jedinci jsou totiž mnohdy součástí různých skupin, a tak je kromě proměnných popisujících tyto jedince (tzv. mikroúroveň) nutné uvažovat i proměnné charakterizující nadřazené skupiny (tzv. makroúroveň). Typickým, v literatuře (Kreft, de Leeuw 1998) velice často uváděným příkladem, jsou studie z oblasti sociologie vzdělání, předmětem jejichž zájmu jsou nejenom jednotliví žáci, ale také konkrétní třídy či školy, jež tito žáci navštěvují. Obdobnou geografickou ukázkou je příslušnost jedinců k obcím či regionům, ve kterých žijí, jejichž prostředí je přirozeným kontextem majícím vliv na nejrůznější jevy a procesy, které stojí v centru pozornosti geografických výzkumů. Obě dvě úrovně proměnných je tak třeba důsledně rozlišovat, mimo jiné také proto, že na každé z nich má smysl zabírat se jinými typy proměnných. Těmito typy proměnných máme na mysli členění podle Lazarsfelda a Menzela (1961, cit. v Hummell, Sodeur 1997, s. 38), kteří vymezují sedm kategorií proměnných. Mezi těchto sedm kategorií patří jak proměnné týkající se jedinců, kterými jsou proměnné absolutní, vztahové, srovnávací a kontextuální, tak také kategorie odkazující na nadřazené skupiny, mezi kterými rozlišujeme proměnné analytické, strukturální a globální.

Prvním typem jsou proměnné absolutní, které jsou ze všech uvedených kategorií nejjednodušší, jelikož se týkají pouze jednoho konkrétního jedince. Vztahu mezi dvěma jedinci, jak naznačuje i samotný název, odpovídají proměnné vztahové. Nepracují však se vztahem napříč úrovněmi, nýbrž pouze se vztahem v rámci jedné úrovně. O porovnávání



napříč úrovněmi tak můžeme hovořit až v souvislosti s proměnnými srovnávacími, které porovnávají, ať již absolutní či vztahové, vlastnosti jedinců s rozložením těchto vlastností v nadřazené skupině. Z vlastností v nadřazené skupině pak vznikají proměnné kontextuální, které daly název i onomu kompletnímu Lazarsfeldově přístupu. Jsou přitom výsledkem desagregace, při které je každému jedinci na nižší úrovni přiřazena hodnota proměnné měřené na vyšší úrovni (Soukup 2006). A právě zavedení těchto kontextuálních proměnných a jejich nemalá role v následných analýzách je zcela klíčovým prvkem víceúrovňově konstruovaných modelů. S jejich pomocí totiž nejsme při kontextuální analýze nuceni volit mezi výzkumem na úrovni jedinců (tzv. mikroanalýza) či na úrovni skupin (tzv. makroanalýza), ale naopak můžeme obě úrovně analyzovat současně v rámci jednoho modelu. Díky této možnosti tak můžeme nejen odpovídat na složitější a komplexnější otázky ohledně analyzovaných vztahů, ale jsme také schopni vyvarovat se chyb a problémů pramenících z takzvaného ekologického usuzování, které je spojeno s izolovanými analýzami kontextuálních dat buď pouze za jedince či pouze za skupiny (Heck a kol. 2012). V prvním z těchto zmíněných případů (mikroanalýza) může přitom docházet k nežádoucímu přehlížení kontextu a ve druhém (makroanalýza) pak hrozí riziko ekologické chyby (Soukup 2006).

Přehlížení kontextových charakteristik prostředí, ve kterém jedinec žije a které ho přirozeně ovlivňují, je spojeno s pojmem atomická chyba (*individualistic fallacy*). Té se dopouštíme v případě, kdy celek chápeme pouze jako součet jeho jednotlivých částí, ale ignorujeme vazby a mechanismy existující mezi těmito jednotlivými částmi, které právě celek dělají celkem (Alker 1969, cit. v Gregory a kol. 2009, s. 174). Jinými slovy odhlížíme od kontextu, výsledkem čehož pak může být nežádoucí ztráta informací z něho plynoucích doprovázená chybným usuzováním z individuálních dat za jednotlivce na celou populaci (Koudelka 2011).

Opakem chyby atomické je podle Gregoryho a kol. (2009) chyba ekologická (*ecological fallacy*), která tedy reprezentuje druhý výše zmíněný problém spojený s izolovanou analýzou kontextuálních dat. Na její existenci poprvé upozornil americký statistik William Robinson v polovině 20. století v souvislosti s analýzou dat pocházejících z patnáctého amerického sčítání z roku 1930, kdy dokázal, že významná ekologická korelace mezi podílem černošských a negramotných obyvatel získaná na úrovni států nemusí být významná na úrovni jedinců (Robinson 1950, cit. v Freedman 1999, s. 1). A právě toto je

situace, ve které může dojít k ekologické chybě, kdy z výsledků analýz dat agregovaných za nadřazené skupiny odvozujeme informace týkající se jedinců, kteří jsou součástí dané skupiny (Gregory a kol. 2009). Jednotlivé konkrétní závěry získané tímto způsobem mohou ovšem být přirozeně zcela správné, čili nelze tvrdit, že tento postup usuzování z celku na části je chybný jako takový. Je ale zapotřebí mít na paměti, že zde existuje značné riziko nesprávného odvození závěrů ohledně mikroúrovně, které je výsledkem takto izolované makroanalýzy v situaci, kdy je mikro- a makroúroveň neoddělitelně provázaná. Tato skutečnost tak ve spojení se situací, kdy je dostupnost agregovaných dat výrazně snazší než získání údajů o jednotlivcích, nahrává zvýšenému nebezpečí ekologické chyby, k jejímuž metodologickému řešení však v nejbližší budoucnosti bohužel zřejmě nedojde (Freedman 1999).

Pro úplnost výše uvedeného popisu jednotlivých typů proměnných je na místě ještě doplnit stručné charakteristiky proměnných týkajících se nadřazených skupin. Proměnné analytické a strukturální jsou podobnými typy proměnných, pro které platí, že jsou konstruovány na vyšších úrovních za pomoci agregování z proměnných měřených na nižších úrovních týkajících se jedinců (Soukup 2006). Rozdíl mezi nimi tkví v tom, že analytické proměnné vznikají z proměnných absolutních, kdežto strukturálním jsou odrazem proměnných vztahových (Hummell, Sodeur 1997). Poslední typ proměnných, kterými jsou proměnné globální, charakterizují Lazarsfeld a Menzel (1969, cit. v Kreft, de Leeuw 1998, s. 9) jako proměnné, které měří kontextové charakteristiky nadřazených skupin přímo, nikoli agregací proměnných týkajících se jedinců měřených na nižších úrovních. Je tedy evidentní, že při takovýchto způsobech vzniku není možné zabývat se těmito typy proměnných na úrovni jednotlivců.

Byť již v této kapitole byla představena řada důvodů, proč je v určitých situacích vhodné analyzovat data víceúrovňově, seznámíme se na následujících stránkách ještě s dalšími, více statistickými argumenty hovořícími pro víceúrovňové modelování.

### **3.2 Statistická motivace aplikace víceúrovňových modelů**

Pro správné uchopení problematiky statistické motivace aplikace víceúrovňových modelů není od věci připomenout to, co již bylo zmíněno výše, tedy že ze statistického hlediska můžeme tyto modely chápat jako zobecnění regresní analýzy. Jedním ze základních

předpokladů této metody je přitom nezávislost jednotlivých pozorování (Hendl 2004). Možná častěji, než si vůbec uvědomujeme, však nastávají situace, kdy není tento požadavek splněn.

Příčinou těchto nežádoucích situací je mnohdy samotný design výzkumu, který není adekvátní vůči následně použitým analytickým metodám. Často například v souvislosti s výzkumy prováděnými v rámci nejruznějších vysokoškolských prací dochází k situaci, kdy výzkumník oslovuje v rámci dotazníkového šetření své spolužáky, kteří chodí do stejné třídy, členy jedné rodiny, sousedy bydlící v jedné lokalitě či obecně řečeno osoby, které jsou vázané k určité entitě, přičemž právě tato vazba na ně s velkou pravděpodobností působí, a ovlivňuje tak jejich odpovědi. To je poté důvodem, proč bychom neměli tyto odpovědi považovat za nezávislá pozorování a data dále zkoumat za pomoci standardních metod regresní analýzy. V takovýchto případech koná výzkumník výše popsaným způsobem většinou spíše z nouze, jelikož není v jeho možnostech provést jakýkoli reprezentativnější výběr, a tak je jeho jedinou volbou (aniž o tom má sám často tušení) onen výběr na základě dostupnosti (Disman 2002).

Avšak ani sofistikovanější metody výběru nejsou zárukou nezávislosti jednotlivých pozorování. V případě výběru víceúrovňového, který je považován za plnohodnotnou náhradu prostého náhodného výběru (Kostecký, Čermák 2003), je tomu spíše naopak. Jedná se totiž o typ výběru, který Disman (2002, s. 107) popisuje jako výběr, který „se provádí ve dvou nebo více krocích. Nejdříve jsou náhodně vybrána určitá přirozená seskupení, a pak teprve jsou náhodně vybíráni jedinci z oněch vybraných seskupení.“ Je tedy evidentní, že jedinci figurující ve finálním výběru jsou součástí určitých nadřazených skupin, přičemž právě tato vazba konkrétních jedinců k určitým nadřazeným skupinám a sdílení společného kontextu je důvodem jejich závislosti (Kreft, de Leeuw 1998). Ilustrační ukázkou může být například finální výběr žáků v pedagogickém výzkumu, ke kterému dojdeme tak, že v prvním kroku náhodně vybereme určité školy, ve druhém náhodně jejich konkrétní třídy a až v rámci třetího kroku jsou opět náhodně vybráni jednotliví žáci. Při geografických víceúrovňových výběrech je možno postupovat srovnatelnou analogií zahrnující například v prvním kroku náhodně vybraný okres, v něm náhodně vybranou obec jako druhý krok a finální náhodný výběr jedinců v dané obci.

Řešení v takovýchto situacích nabízí právě víceúrovňové modely, pro něž není tato hierarchická struktura dat překážkou. Víceúrovňové modely jsou naopak konstruovány právě s cílem zachytit tuto strukturu prostřednictvím více úrovní, které, jak už samotné jejich označení napovídá, zahrnují. V praxi tak nemodelujeme pouze jednu úroveň, nýbrž přistupujeme k tvorbě modelů pro každou relevantní úroveň a propojujeme je v rámci jednoho metodologického postupu. Základem tohoto přístupu je v podstatě provádění dílčích regresních analýz pro jednotlivé skupiny. Tyto dílčí regresní analýzy jsou však spojeny s odhadem určitého počtu parametrů, což v případě většího počtu skupin bude znamenat také velký celkový počet odhadnutých parametrů, který je limitem přesnosti celého modelu. Víceúrovňové modelování si naopak vystačí s menším počtem takovýchto parametrů, které jsou zárukou právě větší statistické přesnosti (Soukup 2006).

Podrobnější diskuze nejrůznějších motivací pro použití víceúrovňově konstruovaných modelů mívá tendence nabudit ve svých potenciálních budoucích uživateli dojem, že je univerzálním všelékem, který je možné aplikovat za každé situace. Tyto modely se podle Soukupa (2006) stávají v zahraničí společenskovědním trendem s cílem použití „moderních“ a sofistikovanějších metod, kterému podléhá celá řada autorů, kteří když neví, co si počít se svou analýzou, zahrnou do ní druhou a třetí úroveň. Díky tomu pro ně i nesmyslná analýza rázem získává nový smysl a to v podstatě pouze díky tomu, že je provedena víceúrovňově, byť nad daty, která pro tento typ analýzy často nejsou vhodná. Pomyslným zdviženým prstem na výstrahu proti tomuto fenoménu by tak měla být následující část konkretizující případy, ve kterých má opravdu smysl přistoupit k víceúrovňově konstruovaným modelům a na co si dát při tomto typu analýzy pozor.

### 3.3 Možnosti a omezení víceúrovňových modelů

Využití možností víceúrovňového modelování má podle Soukupa (2006) smysl v případech, kdy má sledovaná proměnná různé hodnoty u různých jedinců a její průměrná úroveň se zároveň liší v určitých větších celcích (skupinách, hierarchiích). Jinými slovy se jedná o situace, kdy sledovaní jedinci přísluší do určitých skupin, mezi kterými existují významné rozdíly. Takovéto případy jsou, jak zmiňuje Soukup (2006), každodenní realitou, kterou lze vyjádřit statisticky prostřednictvím hodnoty vnitrotřídního koeficientu korelace (ICC). Podrobnějšímu představení tohoto koeficientu včetně způsobu jeho výpočtu a interpretace výsledných hodnot bude věnován prostor v následující části práce

popisující detailní postup konstrukce víceúrovňových modelů. Zde zmíníme pouze skutečnost, že díky konceptu rozkladu rozptylu nám ICC říká, „nakolik se liší hodnoty proměnné na úrovni jednotlivců a nakolik na vyšších úrovních“ (Soukup 2006, s. 994). Ukazuje-li tedy výsledná hodnota na významnou odlišnost mezi oběma úrovněmi, pak má opravdu smysl přistoupit ke konstrukci víceúrovňového modelu.

Kolik úrovní má ale smysl v konkrétní analýze konstruovat? Při veškerých dosavadních diskuzích jednotlivých témat souvisejících s víceúrovňovými modely zůstávala tato otázka, tedy co konkrétně představuje předpona „více“ u víceúrovňových modelů, stranou. V běžné praxi jasně převažují modely pracující se dvěma úrovněmi, tedy takzvané dvouúrovňové. Obecně však lze modelovat i větší množství úrovní, nicméně s jejich rostoucím počtem dochází na jedné straně k nárůstu složitosti samotného modelu a na straně druhé naopak klesá stabilita řešení nalezených takovýchto modely (Snijders, Bosker 2012). Kromě toho také rostou požadavky na velikost výběrového souboru, kdy je zapotřebí mít k dispozici na každé jednotlivé úrovni dostatečný počet pozorování. To je v konečném důsledku právě tím hlavním limitem, který modelování většího počtu úrovní ve skutečnosti skoro neumožňuje (Hox 2010). Existují případy speciálních studií s mnoha tisíci pozorováními, u nichž je možné zavést do modelu ještě třetí úroveň, což je již však opravdu maximální počet úrovní, se kterým se můžeme ve víceúrovňových modelech setkat.

Vyjma těchto standardních modelů, a to ať již se dvěma či třemi úrovněmi, má smysl konstruovat modifikované varianty víceúrovňových modelů také v dalších, na první pohled možná poněkud méně obvyklých situacích, pro něž představuje obecný koncept zavedení více úrovní efektivní možnost jejich řešení. Takovýchto modifikovanými variantami jsou modely růstu, metaanalýza či modely se smíšenými hierarchiemi (Soukup 2006).

Modely růstu jsou víceúrovňovým východiskem analýzy longitudinálních dat, jejímž vstupem jsou hodnoty určité vlastnosti změřené ve více časových okamžicích u různých jedinců, přičemž ona konkrétní měření můžeme v rámci modelu považovat za první úroveň a samotné jedince pak za druhou úroveň. Výsledky modelů růstu poté, obdobně jako standardní víceúrovňové modely, hovoří nejenom o jednotlivých úrovních, ale také o vztazích mezi nimi. Závěrem je tedy jednak zjištění, zda obecně dochází k nárůstu či poklesu hodnot sledované vlastnosti a dále také zda se jedinci liší v růstových křivkách

mezi sebou. Výhodou této metody je skutečnost, že situace, kdy některá pozorování pro určité časové okamžiky u některých jedinců chybí, není pro její náležité použití překážkou, a dokonce ani nevadí, pocházejí-li hodnoty pro různé jedince z měření v různých časových okamžicích, což je zejména vzhledem k problematické práci s chybějícími hodnotami a jejich náhradami ve standardních metodách regresní analýzy výhoda více než značná (Soukup 2006).

Metaanalýza je v jistém ohledu odlišným typem analýzy, která sama nepracuje s žádnými primárními daty. Jejím cílem je syntéza číselných výsledků již publikovaných studií těchto primárních dat (Glass 1976), k jehož naplnění může opět přispět právě zkonstruování víceúrovňového modelu. Budeme-li považovat data za jedince z jednotlivých studií za první úroveň, pak mohou být samotné studie, analogicky jako nadřazené skupiny ve standardních víceúrovňových modelech, chápány jako druhá úroveň. Aplikace tohoto pohledu nám poté umožní jednak najít společný výsledek všech nalezených studií a také odhalit příčiny rozdílů mezi studiemi (Hox 2010).

Posledním z příkladů modifikovaných variant víceúrovňových modelů, který si zde blíže představíme, jsou modely se smíšenými hierarchiemi, jejichž teoretické pozadí je v porovnání s předchozími příklady o poznání složitější a komplikovanější. Tyto komplikace souvisí s problematikou vícenásobné příslušnosti jedinců do různých skupin, která je souhrnným označením pro situace, kdy neexistuje pouze jedna jediná skupina, která má na konkrétního jedince vliv. S ohledem na již několikrát zdůrazňovanou velikost a složitost přirozeného systému v sociálních vědách tato skutečnost jen stěží někoho překvapí, stále však znamená pro statistickou analýzu komplikace tím spíše, že jednotlivé skupiny často nejsou ani samy o sobě hierarchicky uspořádané. Jen stěží posoudíme, zda má na studijní výsledky konkrétního žáka větší vliv třída, potažmo učitel, který v ní vede vyučování, či rodinné prostředí, ze kterého daný žák pochází a kde se věnuje domácí přípravě. Stejně tak například v geografickém výzkumu neexistuje jednoznačná odpověď na to, zda mají v regionální hierarchii vyšší postavení regiony vymezené na základě dojížděky za prací anebo administrativní regiony. Avšak ani takovéto komplikace nejsou překážkou, se kterou by si modely se smíšenými hierarchiemi neporadily, neboť jejich konstrukce jim umožňuje reflektovat a zohledňovat onu příslušnost jedince do několika různých skupin. Ten tak není v průběhu analýzy zařazen pouze do jedné skupiny, která ho

nejvíce ovlivňuje, ale naopak do každé jednotlivé skupiny, o níž se v rámci analýzy domníváme, že může mít významný vliv (Soukup 2006).

Některá z úskalí, na která je třeba si při víceúrovňovém modelování dát pozor, již byla krátce zmíněna a odhalena, nyní však budou potenciální nebezpečí diskutována podrobně. Jedná se zejména o již zmíněnou problematiku výběru a jeho velikosti a dále problémy související s příslušností jedinců do nadřazených skupin. Těmi jsou konkrétně vícenásobná příslušnost jedinců do několika různých skupin, kterou jsme již též zmínili, a dále potíže související s ohraničením jednotlivých skupin a s pohybem jedinců mezi skupinami (Snijders, Bosker 2012).

Nezbytná minimální velikost výběrového souboru není v případě víceúrovňového modelování pouze drobným úskalím, nýbrž spíše zcela zásadním limitujícím faktorem, který je pravděpodobně jedním z nejvýznamnějších omezení masovějšího rozšíření této metody. Ta je v důsledku svého založení na podobných principech jako lineární regresní analýza vystavena celé řadě obdobných předpokladů, které musí studovaná data splňovat (Voda 2013), tudíž ani samotná velikost souboru analyzovaných dat se nevyhne určitým pravidlům. Jedno z nejjednodušeji formulovaných pravidel říká, že při konstrukci regresního potažmo víceúrovňově stavěného modelu by na jednu proměnnou mělo připadat deset pozorování (raději 100 jedinců), což by v případě víceúrovňových modelů mělo platit i na vyšších úrovních (jedna proměnná - deset nadřazených skupin). Je evidentní, že nároky na minimální počet jedinců studovaných za pomoci víceúrovňově konstruovaných metod rostou velmi rychle a již například situace, kdy hodnotíme tři proměnné, znamená analogicky také 30 nadřazených skupin a v každé z nich 30 jedinců, čili dohromady 900 jednotek, které bychom měli mít pro takovou to analýzu k dispozici (Hox 2010). 30 skupin je také považováno za ideální počet nadřazených jednotek, ovšem jako opravdu minimální je jejich počet stanoven na 20, přičemž výjimkou však nejsou ani studie, ve kterých data náleží do pouhých 14 skupin (Voda 2013).

Potíže s ohraničením jednotlivých skupin vycházejí stejně jako i oba dva další výše jmenované problémy ze společenské povahy analyzovaných dat. Takovýto typ dat lze v některých případech jen stěží přesně klasifikovat a zařadit do ostře ohraničených skupin, které nerespektují jejich spojitý charakter. Při víceúrovňových analýzách se však bez takovýchto přesně vymezených skupin neobejdeme, proto je nezbytné u veškerých, tudíž

i spojitých dat, ostré hranice vytvořit. Při interpretaci výsledků je poté potřeba mít na paměti, že data byla zpracována takovýmto způsobem.

Situaci, se kterou se relativně běžně setkáváme, představuje pohyb jedinců mezi skupinami. Jejich příslušnost do nadřazených skupin totiž není stálou a neměnnou záležitostí, což může hrát zkreslující a limitující roli. Příkladem mohou být situace, kdy například žák přestoupí do jiné třídy, jedinec migruje či dojde ke změně příslušnosti obce k danému regionu vlivem úpravy jeho spádových hranic. Rázem se tak v těchto případech mění kontext, který konkrétní žáky, jedince či obce ovlivňuje.

Uvědomění si těchto nebezpečí a adekvátní příprava analyzovaných dat je tedy nezbytnou podmínkou pro správné vybudování víceúrovňových modelů, jehož podrobnější diskuze je hlavní náplní následující části.

### 3.4 Konstrukce víceúrovňových modelů

Pro proces konstrukce víceúrovňových modelů je příznačný určitý postupný průběh, v jehož rámci je souhrnný model budován krok za krokem po jednotlivých úrovních. Modelování vztahů na první úrovni zpravidla předchází zavedení dalších proměnných na druhé úrovni (případně dalších následujících úrovních). Přestože se přesná podoba výsledného souhrnného modelu může v konkrétních případech lišit, můžeme v průběhu jeho konstrukčního procesu vymezit pět samostatných kroků a jim odpovídajících modelů. Prvním z nich je model prázdný obsahující pouze závisle proměnnou a identifikaci nadřazených skupin (model 0). Po jeho konstrukci přichází na řadu přidání individuálních nezávisle proměnných (model 1) a umožnění rozdílných hodnot sklonů přímky v různých nadřazených skupinách (model 2). Poté je již souhrnný model budován na dalších úrovních, a to nejprve prostřednictvím zavedení skupinových nezávisle proměnných (model 3) a následně také přidání meziúrovňových interakcí (model 4) (Voda 2013).

Konstrukce nulového, někdy též nazývaného prázdného modelu obsahujícího tedy, jak již bylo uvedeno výše, pouze závisle proměnnou a identifikaci skupin, je nezbytným počátečním krokem budování víceúrovňového modelu. Právě díky němu jsme totiž schopni učinit opodstatněné rozhodnutí, zda má opravdu smysl k jeho konstrukci přistoupit. Kýženým opodstatněním je v tomto případě vnitrotřídní koeficient korelace



(*intra-class correlation coefficient*, ICC), který byl již v krátkosti zmíněn a diskutován v rámci předchozí části, a o němž tedy víme, že nám říká „nakolik se liší hodnoty proměnné na úrovni jednotlivců a nakolik na vyšších úrovních“ (Soukup 2006, s. 994). Tuto odlišnost můžeme na základě konceptu rozkladu celkového rozptylu na rozptyl vnitroskupinový a meziskupinový číselně vyjádřit pomocí vzorce:

$$ICC = \frac{\sigma_u^2}{\sigma_e^2 + \sigma_u^2},$$

kde  $\sigma_u^2$  je rozptyl na úrovni jedinců a  $\sigma_e^2$  je rozptyl na úrovni nadřazených skupin (Gelman, Hill 2007). Výsledná hodnota pak po vynásobení 100 udává, kolik procent z celkového rozptylu připadá na rozdíly mezi nadřazenými skupinami. Jsou-li tyto rozdíly malé (hodnota ICC je nižší než 5 %), nemá použití metody víceúrovňového modelování příliš velký smysl, neboť rozdíly mezi konkrétními jedinci jsou velmi málo závislé na rozdílech mezi nadřazenými skupinami (Voda 2013). Stejně tak platí, že pokud vyjde hodnota ICC vysoká, pak celkové rozdíly pocházejí zejména z rozdílů mezi nadřazenými skupinami a vhodným analytickým postupem je spíše standardní regresní analýza se skupinovými daty (Gelman, Hill 2007). Podle Nezleka (2008) je nicméně použití víceúrovňově konstruovaných modelů vhodné ve všech případech, kdy mají analyzovaná data hierarchickou strukturu, třebaže hodnota ICC nevyjde vysoká. V každém z těchto případů pak tedy můžeme postoupit k budování prvního potažmo druhého modelu.

Pro úplnost popisu nulového modelu je na místě ještě doplnit jeho formální zápis:

$$Y_{ij} = y_{00} + u_{0j} + e_{ij} \text{ (model 0),}$$

kde  $Y_{ij}$  je symbol pro hodnoty závisle proměnné,  $y_{00}$  je odhadem střední hodnoty závisle proměnné v celém základním souboru,  $u$  a  $e$  jsou náhodné chyby na druhé, resp. první úrovni, index  $i$  udává pořadí jedince v příslušné nadřazené skupině a index  $j$  je pořadové číslo nadřazené skupiny (Soukup 2006).

První i druhý model primárně vznikají přidáním individuálních nezávisle proměnných do nulového modelu. Tento krok nabízí ve své podstatě dva možné přístupy, které jsou reflektovány právě skrze ony dva modely. První možností je definování nových proměnných jako tzv. fixního efektu, druhou poté jako tzv. náhodného efektu. Soukup (2006) ve svém příspěvku formuluje alespoň přibližné pravidlo, na základě kterého

se ve věci definice nových proměnných můžeme rozhodnout, a to že pokud jsou hodnoty nové proměnné obsaženy ve výběrových datech všechny (například jsou zastoupeni muži i ženy a víc kategorií pohlaví neznáme), užijeme fixní efekty. Pokud tomu tak není, například máme-li ve výběrových datech jen osoby z náhodně vybraných lokalit a chceme zobecňovat na všechny lokality, nezbyvá nám nic jiného než definovat vysvětlující proměnnou jako náhodnou.

Užití fixních efektů je tedy zohledněno prvním modelem, někdy též nazývaným modelem s náhodnou konstantou, který slouží jednak ke zjištění, zda vůbec mají očekávané vysvětlující proměnné na individuální úrovni vliv na závisle proměnnou a zda se případně liší regresní konstanty u jednotlivých nadřazených skupin (Voda 2013), jak je schematicky znázorněno na obrázku 1. Formální zápis prvního modelu pak vypadá následovně:

$$Y_{ij} = y_{00} + y_{01} * X_{ij} + u_{0j} + e_{ij} \text{ (model 1),}$$

kde oproti nulovému modelu (model 0) přibyl člen udávající průměrnou směrnici regresních přímk ve všech nadřazených skupinách v celém základním souboru ( $y_{01}$ ) a symbol pro hodnoty nezávisle proměnné ( $X_{ij}$ ) (Soukup 2006, Rychtaříková 2015).

Jako náhodná je vysvětlující proměnná definovaná druhým modelem, někdy též nazývaným modelem s náhodnou konstantou i směrnicí, jehož význam stojí na stejných základech jako význam modelu předchozího. Kromě odlišných regresních konstant napříč jednotlivými nadřazenými skupinami však mohou ve druhém modelu, jak je schematicky znázorněno na obrázku 2, figurovat i odlišné směrnice. Náhodnost vysvětlující proměnné totiž v tomto případě způsobuje její rozdílný efekt v rámci jednotlivých nadřazených skupin. Tato modifikace se samozřejmě projeví i na formálním zápisu, který má pro druhý model tuto podobu:

$$Y_{ij} = y_{00} + y_{01} * X_{ij} + u_{1j} * X_{ij} + u_{0j} + e_{ij} \text{ (model 2),}$$

kde oproti prvnímu modelu (model 1) přibyl člen udávající náhodnou chybu směrnic regresních přímk v jednotlivých nadřazených skupinách v celém základním souboru ( $u_{1j}$ ) (Soukup 2006, Rychtaříková 2015).

Třetí a čtvrtý model již dokonale vystihují podstatu víceúrovňovosti modelu. Ta je ztvárněna budováním jeho dalších úrovní, jehož cílem je vysvětlit rozptyl na druhé úrovni, a to zejména rozptyl náhodných efektů první úrovně za pomoci kontextuálních proměnných, které přísluší druhé úrovni (Soukup 2006). Z tohoto důvodu tak jsou do třetího modelu, někdy též nazývaného modelu s náhodnou konstantou a proměnnými na první i druhé úrovni, zavedeny skupinové nezávisle proměnné reprezentující právě přirozený kontext. Formální zápis třetího modelu pak vypadá následovně:

$$Y_{ij} = y_{00} + y_{01} * X_{ij} + y_{10} * Z_j + u_{1j} * X_{ij} + u_{0j} + e_{ij} \text{ (model 3),}$$

kde oproti druhému modelu (model 2) přibyl člen udávající závislost velikosti regresní konstanty na skupinové nezávisle proměnné ( $y_{10}$ ) a symbol pro hodnoty skupinové nezávisle proměnné ( $Z_j$ ) (Soukup 2006).

Čtvrtý model, někdy též nazývaný model s náhodnou konstantou i směrnici a proměnnými na první i druhé úrovni, má v porovnání s modelem třetím analogicky stejnou přidanou hodnotu jako měl druhý model oproti prvnímu. I zde jsou modelovány směrnice, které mohou náhodně variovat a navzájem se tak odlišovat. Jedná se přitom o odlišnosti projevující se vlivem synergického působení proměnných na první a druhé úrovni. Tyto interakce se samozřejmě projeví i na formálním zápisu, který má pro čtvrtý model tuto podobu:

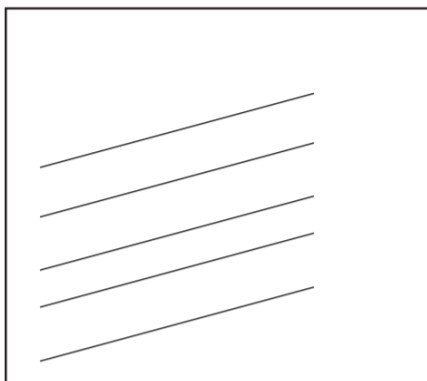
$$Y_{ij} = y_{00} + y_{01} * X_{ij} + y_{10} * Z_j + y_{11} * X_{ij}Z_j + u_{1j} * X_{ij} + u_{0j} + e_{ij} \text{ (model 4),}$$

kde oproti třetímu modelu (model 3) přibyl člen udávající závislost velikosti regresní konstanty na interakcích mezi hodnotami nezávisle proměnné a skupinové nezávisle proměnné ( $y_{11}$ ) a symbol pro interakce mezi oběma úrovněmi ( $X_{ij}Z_j$ ) (Soukup 2006, Rychtaříková 2015).

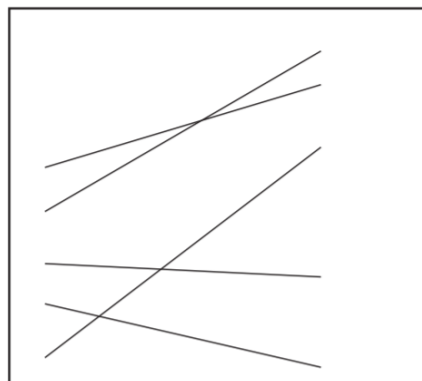
Uvedený formální zápis čtvrtého modelu je zároveň matematickou formulací plného víceúrovňového lineárního modelu, pomocí kterého jsou tedy hodnoty závisle proměnné dány součtem celkového průměru závisle proměnné, efektu individuální nezávisle proměnné, efektu skupinové nezávisle proměnné, efektu interakce individuální a skupinové a proměnné, efektu interakce chyby na skupinové úrovni a individuální nezávisle proměnné a dvou chybových složek (Voda 2013). Jedná se přitom o zápis

nejpoužívanějšího víceúrovňového lineárního modelu se spojitou závisle proměnou. I když je modelování jiných typů proměnných, jakými jsou například proměnné binární, nominální či ordinální, prostřednictvím víceúrovňově konstruované analýzy teoreticky samozřejmě také proveditelné, je tento postup v praxi značně marginální.

**Obrázek 1:** Regresní přímky s odlišnými konstantami



**Obrázek 2:** Regresní přímky s odlišnými konstantami i směnicemi



**Zdroj:** Rychtaříková 2015, úprava autorky

**Zdroj:** Rychtaříková 2015, úprava autorky

Důležitou součástí procesu konstrukce víceúrovňových modelů je jeho softwarové zajištění. Možnosti v podobě komerčně nabízených a široce využívaných statistických softwarů jakými jsou například SAS, STATA či SPSS již byly zmíněny. Vyvinuto však bylo i mnoho programů zaměřujících se speciálně na víceúrovňové modely. Mezi ně patří například HLM, BMDP5-V, MlwiN, Mplus, MIXOR či MIXREG (Kreft, de Leeuw 1998, Rychtaříková 2015). V případě analýzy v rámci této práce však byl upřednostněn statistický software SPSS.

Moderní výpočetní technika ilustrovaná tedy i těmito softwary je při konstrukci víceúrovňových modelů velkým pomocníkem. Její progresivní vývoj bezesporu významně přispívá ke snazšímu a kvalitnějšímu přechodu mezi teoretickými aspekty zkoumané problematiky a její praktickou aplikací. Přestože je důležitost teoretických aspektů víceúrovňového modelování značná, čemuž odpovídá i prostor jim věnovaný v rámci této práce, je to právě aplikační část jednotlivých geografických, výzkumných studií, která má z pohledu výzkumu reality klíčový význam. Právě proto je podrobnější přehled a bližší představení konkrétních příkladů celé řady takovýchto aplikací víceúrovňového modelování v empirickém výzkumu náplní následující kapitoly.

## Kapitola 4

### Příklady aplikací víceúrovňových modelů

Sestavení vyčerpávajícího přehledu příkladů aplikací víceúrovňově konstruovaných modelů je úkolem velice složitým, ne-li nespílnitelným, jelikož situací, ve kterých je využití víceúrovňových modelů žádoucí a přínosné, je opravdu mnoho. To ve své knize potvrzují i Kreft a de Leeuw (1998, s. 1), kteří tvrdí, že „jakmile víte, že existují hierarchie, vidíte je všude“. Množství těchto opodstatněných aplikací je bohužel ještě obohacováno celou řadou nesmyslných analýz, jejichž záměrem není správná aplikace víceúrovňových modelů, nýbrž pouze následování společenskovedního metodologického trendu s cílem použití „moderních“ a sofistikovanějších metod, který jsme zmiňovali výše. Níže uváděný přehled je tak spíše přehledem nejrozumnějších oblastí, ve kterých víceúrovňové modely nacházejí své uplatnění.

První místo si bezesporu nezaslouží nikdo jiný než pedagogika a sociologie vzdělávání, které jsou ve své podstatě průkopníky masivnějšího využívání víceúrovňových modelů. Donald Rubin (1980, cit. v Bock 1989) se již na začátku 80. let 20. století snažil za pomoci víceúrovňově konstruovaného modelu podrobněji popsat závislost mezi studijním průměrem jednotlivých studentů prvních ročníků právnických oborů vysokých škol a jejich studijním průměrem na střední škole. Na základě údajů z přijímacích řízení 82 právnických škol došel k závěru, že víceúrovňová analýza je opravdu schopna lépe reflektovat tento vztah, a přesněji tak předpovědět studijní průměr studentů prvních ročníků těchto škol na základě jejich středoškolského studijního průměru. O tři roky později se pak ještě s Braunem, Jonesem a Thayerem (1983, cit. v Bock 1989) snažil najít odpověď na otázku, zda má příslušnost k menšině vliv na vztah studijního průměru studentů prvních ročníků ekonomicky zaměřených vysokých škol a výsledků jejich přijímacích testů do

vysokoškolského studia. I zde se použití víceúrovňové analýzy ukázalo jako relevantní, neboť pro příslušníky minorit zde existoval jiný vztah než pro bělošské obyvatelstvo. Tradičním oborem, ve kterém se víceúrovňově konstruované modely právem těší velké oblibě, je oblast zdraví a epidemiologie. Vliv faktorů vnějšího prostředí na individuální vztah, potažmo náchylnost jedinců k určitým problémům či onemocněním, je v rámci epidemiologických studií tématem vskutku obvyklým. Škála konkrétních zaměření sahá od nízké porodní hmotnosti (O'Campo a kol. 1997; Roberts 1997) přes kojeneckou úmrtnost (Matteson, Burr, Marshall, 1998), kouření (Kleinschmidt, Hills, Elliott, 1995; Duncan, Jones, Moon, 1999), obezitu (Rundle a kol. 2007) až po nejružnější příčiny úmrtnosti v dospělém věku, jakými jsou například vnější příčiny (Cubbin, LeClere, Smith 2000) či kardiovaskulární onemocnění (Diez-Roux a kol. 1997; Davey Smith a kol. 1998; Sundquist, Malmstrom, Johansson 1999).

Podstatnou je však v této souvislosti skutečnost, že kontextuální faktory vnějšího prostředí obvykle reprezentují určité vymezené oblasti či regiony, čili se ve své podstatě jedná o výzkumy svým způsobem geografické. Toto je přitom charakteristický rys celé řady víceúrovňových analýz, které v rámci kontextu chápaného jako určitá geografická jednotka studují nejružnější konkrétní témata. Kromě výše zmíněných se jedná například o problematiku hodnocení vlastního zdraví (Kawachi, Kennedy, Glass 1999; Subramanian, Kim, Kawachi 2002; Kim, Subramanian, Kawachi 2006), o faktory ovlivňující porodnost (Billy, Moore 1992; Dribe, Juárez, Scalone 2015), nezaměstnanost (Brown, Session 1997; Gould, Fieldhouse 1997; Khattab 2006; Arntz, Wilke 2009; Baum, Bill, Mitchell 2009; Baum, Mitchell 2010) či migraci (Yang, Guo 1999; Ezra, Kiros 2006; Kulu, Billari 2010). Výjimkou však také není ani setkání s víceúrovňově konstruovanými modely v sociologii (Hsu, Chang, Yip 2016), ekonomii (Lederman 2010), psychologii (Schyns, Van Veldhoven 2010; Ling, Guo, Cheng 2018) či kriminologii (Sampson, Raudenbush, Earls 1997; Zhang, Messner, Liu 2007; Moura, Silveira-Neto 2015),

České prostředí v podstatě kopíruje výše nastíněný světový model používání víceúrovňových analýz, a tak nás jejich oborová aplikace zřejmě nepřekvapí. Odlišná je spíše četnost využití víceúrovňových modelů, která je u nás nesrovnatelně menší než v zahraničí. Řadu studií (Soukup 2006; Pecha 2009; Voda 2013) lze považovat za spíše úvodní práce věnující se komplexně problematice metody víceúrovňového modelování, a o jejich aplikačních částech tak hovoříme spíše jako o didaktických

příkladech použití metody než o kompletních, empiricky zaměřených vědeckých studiích. Ani tak však není o pestré tematice nouze a najdeme u nás významné příklady aplikací jak z oblasti pedagogiky a sociologie vzdělávání (Soukup 2006; Pecha 2009; Straková 2010), tak také z obecné sociologie (Nový 2015; Kudrnáč 2017) či geografie (Hamplová 2006; Hamplová, Řeháková 2006; Spilková, Dzúrová, Pikhart 2011; Voda 2013; Rychtaříková 2015).

První z uvedených geografických článků od Hamplové (2006) se zaměřuje na mapování vztahu mezi rodinným stavem a životní spokojeností v jedenadvaceti evropských zemích. Pozitivní vliv manželství na životní spokojenost z výsledků vyplývá jednoznačně, avšak síla tohoto vlivu je mezi jednotlivými státy odlišná.

Příspěvek Hamplové a Řehákové (2006) se zabývá analýzou vztahu mezi mimomanželskou plodností na našem území po roce 1989, individuálními charakteristikami jednotlivých žen a socioekonomickou situací konkrétních regionů. Za klíčovou přitom považuje odpověď na otázku, „jak ekonomická situace regionu a z ní vyplývající struktura příležitostí ovlivňuje šance, že se žena stane svobodnou (či obecně neprovdanou) matkou, případně či šance jsou nejvíce ovlivněny“ (Hamplová, Řeháková 2006, s. 28). Vliv socioekonomické situace na pravděpodobnost, že žena porodí dítě neprovdaná, tato studie prokázala a upozornila také na jeho větší sílu u skupin mladých a málo vzdělaných žen.

Stejně jako Hamplová (2006) se tématem životní spokojenosti v rámci svého výzkumu zabíral také Voda (2013). Porovnávanými oblastmi však v jeho případě nebyly evropské státy, ale obce ležící ve východních Čechách. U nich za pomoci víceúrovňově konstruované analýzy zjišťoval, které individuální a kontextuální faktory mají vliv na občanskou spokojenost, zda se liší vliv individuálních faktorů mezi jednotlivými obcemi a co případně tuto odlišnost způsobuje. Na základě provedené analýzy 916 individuálních dotazníků bylo zjištěno, že úroveň spokojenosti občanů se v jednotlivých obcích liší, přičemž obecně platí, že s vyšším příjmem a vyšším věkem občanů jejich spokojenost roste. Náhodné efekty však ukázaly, že efekt věku a příjmu se v jednotlivých obcích odlišuje, což značí notný vliv kontextu, ve kterém konkrétní jedinci žijí.

Studie Spilkové, Dzúrové a Pikharta (2011) je zástupcem tradiční výzkumné oblasti zdraví a epidemiologie. Svou pozornost upíná k v současnosti výraznému celosvětovému problému spojenému se zdravotním stavem, kterým je kouření. Zkoumá především nerovnosti ve vztahu ke kouření mezi českou populací, v rámci kterých odhaluje pomocí víceúrovňově konstruované analýzy řadu pravidelností. Na individuální úrovni identifikuje jako rizikovější skupiny s větší šancí na to být kuřákem muže, osoby nižšího věku a s nižším vzděláním. Významně více kuřáků je dle této studie dále také mezi osobami rozvedenými a nezaměstnanými. Co se týče úrovně kontextuální potažmo regionální, vyšší šanci na to být kuřákem vykazovaly pro své obyvatele oblasti s vyšší mírou nezaměstnanosti a s vyšším podílem cizinců.

Aktuální práce Rychtaříkové (2015) se dotýká v současnosti hojně diskutované oblasti zdraví a délky života, která patří k obsahově bohatší a složitější problematice. Proto byla před samotnou dvouúrovňovou analýzou postojů obyvatel ke zdraví, životu a stáří ve třiceti evropských zemích, provedena deskriptivní analýza délky života a zdravotního stavu ve věku 65+ let na celém sledovaném území, která byla vhodným předstupněm mapujícím studovanou oblast. Následná analýza pracovala se třemi modely vyjadřujícími vztah mezi postoji obyvatel a pohlavím, věkem, vzděláním a rodinným stavem na první úrovni a příslušností k dané zemi na druhé úrovni. V rámci prvního modelu byla podrobněji analyzována spokojenost obyvatel se zdravím, v rámci druhého spokojenost se životem a třetí model více souvisel s kulturními zvyklostmi označování věku stáří. Nejvýznamnější regionální vlivy byly prokázány u druhého modelu, kde hodnota ICC překročila deset procent. Pozitivní vliv na spokojenost se životem prokázalo prostředí skandinávských zemí, zatímco na opačné straně spektra se nacházelo například Maďarsko, Bulharsko či Lotyšsko.

Víceúrovňově konstruovaná analýza nezaměstnanosti tento přehled příkladů aplikací víceúrovňových modelů u nás v tuto chvíli nedoplní, jelikož k jejímu provedení dosud v českém prostředí nedošlo. Z tohoto důvodu je tak hlavním zdrojem informací pro následující kapitulu diskutující právě problematiku víceúrovňových modelů a nezaměstnanosti výše citovaná zahraniční literatura.



## Kapitola 5

### Víceúrovňové modely a nezaměstnanost

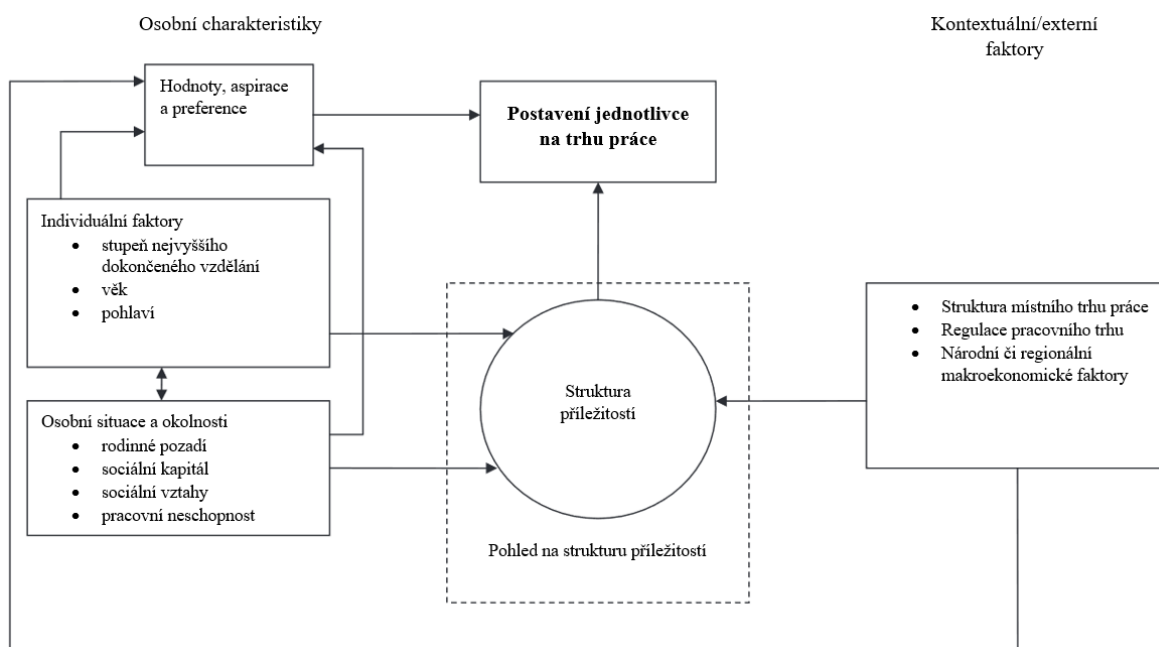
Využití víceúrovňového modelu jako vhodného nástroje studia nezaměstnanosti je v zahraničí uskutečňovanou praxí již od 90. let 20. století. Řada studií (Brown, Session 1997; Gould, Fieldhouse 1997; Khattab 2006; Arntz, Wilke 2009; Baum, Bill, Mitchell 2009; Baum, Mitchell 2010) právě díky tomuto typu analýzy prokázala existenci vlivu jak individuálních, tak regionálních determinantů nezaměstnanosti, které byly modelovány v rámci více úrovní. Analýze vztahu mezi nezaměstnaností a determinanty individuálními přitom patřila zpravidla první úroveň, kdežto modelování souvislostí nezaměstnanosti a determinantů regionálních bylo hlavním smyslem druhé úrovně.

Individuální determinanty, osobní charakteristiky (*personal characteristics*) či obecně faktory interní, jak bývají nezávislé proměnné zařazené do první úrovně víceúrovňových modelů také nazývány, souvisejí se vztahem osobních charakteristik konkrétních jedinců a jejich příležitostí na trhu práce, potažmo jejich nezaměstnaností. Můžeme je ještě dále dělit na faktory čistě individuální (*individual factors*), mezi které patří například stupeň nejvyššího dokončeného vzdělání, věk či pohlaví a faktory vztahující se spíše k osobní situaci a okolnostem (*personal circumstances*) (Baum, Mitchell 2009). Tyto faktory obvykle reprezentuje rodinné pozadí, sociální kapitál, sociální vztahy, případně určitá pracovní (ne)schopnost konkrétního jedince, tedy proměnné související spíše se sociálním a rodinným zázemím.

Hovoříme-li o determinantech regionálních, respektive obecně faktorech externích ve smyslu terminologie užitě v předchozím odstavci, máme tím na mysli zejména proměnné popisující prostředí a v něm působící instituce. Konkrétně se může jednat

například o strukturu místního trhu práce, regulace pracovního trhu nebo národní či regionální makroekonomické faktory. V důsledku spolupůsobení obou těchto skupin faktorů, které je schematicky znázorněno na obrázku 3, je tedy riziko nezaměstnanosti u konkrétních jedinců ovlivněno jak determinanty individuálními, tak také regionálními, jak již bylo ostatně zmíněno výše. Regionální determinanty, jež tvoří kontext celé této problematiky, však působí na riziko nezaměstnanosti naprosto nezávisle na osobních charakteristikách konkrétních jedinců. Hypotetické riziko nezaměstnanosti dvou jedinců s naprosto identickými osobními charakteristikami tak může být odlišné, budou-li se tito jedinci nacházet v prostoru geograficky různých trhů práce. To stejné pak platí i v opačné situaci dvou osob nacházejících se na stejném pracovním trhu, které se však navzájem liší svými osobními charakteristikami.

**Obrázek 3: Model rizika nezaměstnanosti**



**Zdroj:** Baum, Mitchell 2009, úprava autorky

Tyto závěry nemají své důležité místo pouze ve sféře vědecké, teoreticky zaměřené, nýbrž také v oblasti aplikační na poli tvorby nejrozličnějších politických rozhodnutí a opatření, kterým víceúrovňové modely mohou přinést zcela nový, komplexnější pohled na věc. Řada sociálních či ekonomických problémů, nezaměstnanost nevyjímaje, se řadí k takzvaným *wicked problems* (Baum, Mitchell 2009), což jsou právě široce podmíněné problémy značně komplexní povahy. Pro jejich řešení, které často vyžaduje spolupráci mnoha různých subjektů, jsou závěry víceúrovňových analýz klíčovým podkladem.

Tvorba takovýchto politicky orientovaných opatření souvisejících s problematikou nezaměstnanosti však není ani v nejmenším ambicí, ke které by směřovala analýza nezaměstnanosti, která je hlavní náplní následujících částí. Jejím cílem je jednak samotná ukázka aplikace metody víceúrovňového modelování v praxi v konkrétním geografickém výzkumu a dále také identifikace a podrobnější studium jednotlivých individuálních a regionálních proměnných souvisejících s nezaměstnaností.

## 5.1 Příprava dat pro analýzu

Datový soubor představující ústřední podklad pro analýzu rizika nezaměstnanosti je tvořen individuálními anonymizovanými daty za Sčítání lidu, domů a bytů 2011 (SLDB 2011). Tento kompletní soubor obsahoval údaje o 10 436 560 osobách, mimo jiné také informaci o jejich regionální příslušnosti vyjádřenou kódem obce obvyklého bydliště. Vzhledem k celkovému zaměření veškerých analýz na problematiku nezaměstnanosti byly z tohoto souboru vyřazeny jednak osoby ekonomicky neaktivní a dále pak také osoby, které sice při sčítání ekonomickou aktivitu deklarovaly, ale věkově však spadaly do kategorií označovaných jako ekonomicky neaktivní, kterými rozumíme kategorie mladší patnácti let a starší šedesáti čtyř let. Po této redukci byl soubor tvořen 4 660 148 osobami.

V tomto momentě přišlo na řadu další zúžení souboru, a to konkrétně podle dostupnosti vybraných ukazatelů. Jednalo se o krok, kterému předcházela samotný výběr oněch ukazatelů, kterými byly ukazatele pravděpodobně ovlivňující hodnotu binární závisle proměnné nezaměstnanost. Tyto ukazatele byly vybrány na základě studia výše uvedené literatury a v ní analyzovaných individuálních determinantů, které jsou schematicky znázorněné v tabulce 1. V případě analýzy v rámci této práce se pak, s ohledem na české prostředí a dostupnost dat, konkrétně jednalo o ukazatele věk, rodinný stav/registrované partnerství, život s partnerem, rodinný vztah, nejvyšší ukončené vzdělání, obor vzdělání, pohlaví, život v domácnosti s alespoň jedním dítětem a náboženská víra osoby.

Základní ukazatele jako například pohlaví či věk nebyly přirozeně proměnnými, jejichž dostupnost by využití co největšího počtu pozorování omezovala. Zásadními ukazateli byly v tomto směru náboženská víra, jejíž uvedení není ve sčítání povinné, či obor vzdělání. Výběrem osob, u kterých tedy byly k dispozici informace o veškerých výše uvedených proměnných, došlo k zeštíhlení souboru na 2 651 922 osob, což však stále ještě nebyl počet

konečný, který by byl podkladem následného analytického procesu. Smyslem poslední redukce datového souboru bylo reflektovat národnost jednotlivých osob a podle ní pak do konečného souboru vybrat pouze ty osoby, které do svých sčítacích listů vyplnily národnost českou, moravskou či slezskou, a to ať již jako jedinou národnost nebo případně jako jakoukoliv součást kombinace při dvojím uvedení národnosti, které SLDB 2011 umožňovalo. Tato národnost zohledňující úprava tak byla, posledním a také řádově nejmenším zúžením datového souboru, po němž bylo součástí jeho konečné podoby 2 232 757 osob, které se staly předmětem provedené analýzy.

**Tabulka 1: Individuální determinanty nezaměstnanosti analyzované v jednotlivých studiích**

Studie	Analyzované individuální determinanty nezaměstnanosti						
	Věk	Rodinný stav	Vzdělání	Pohlaví	Zaměstnání	Rasa	Ostatní <sup>1</sup>
Brown, Session 1997	✓	✓	✓	✓	✓	✓	příjem, děti
Gould, Fieldhouse 1997	✓	✓*		**	✓	✓	odvětví ekonomické činnosti
Khattab 2006	✓	✓*	✓	**	✓	✓	imigrační status
Arntz, Wilke 2009	✓	✓*	✓	✓			příjem, děti
Baum, Bill, Mitchell 2009	✓	✓*	✓	✓			pracovní (ne)schopnost
Baum, Mitchell 2010	✓	✓*	✓	✓			rodinné pozadí

**Poznámky:** ✓\* analýza rodinného stavu prostřednictvím binární proměnné s variantami ženatý/vdaná a ostatní; \*\* studie zaměřující se pouze na analýzu mužské nezaměstnanosti; <sup>1</sup>nejedná se o vyčerpávající přehled, pouze ukázka některých z dalších analyzovaných proměnných

**Zdroj:** Brown, Session 1997; Gould, Fieldhouse 1997; Khattab 2006; Arntz, Wilke 2009; Baum, Bill, Mitchell 2009; Baum, Mitchell 2010, úprava autorky

Na tomto místě je důležité poznamenat, že přestože je tento datový soubor v dalším průběhu analytického procesu označován jako soubor základní, ze kterého vycházely například následné reprezentativní výběrové soubory, sám o sobě svou reprezentativnost bohužel zaručit nemůže. Zdrojem této nereprezentativnosti je právě výběr podle dostupnosti vybraných ukazatelů, na základě kterého pravděpodobně došlo k nenáhodnému zmenšení souboru. Analýza provedená pouze za osoby, které tedy ve sčítacích listech vyplnily kolonky u veškerých výše uvedených proměnných, zvyšuje na druhou stranu hodnotu takového rozboru v tom smyslu, že odpovědi těchto osob budou s větší pravděpodobností více založené na pravdě než údaje o osobách, které vyplnily pouze některé informace.

Základní soubor měl kromě své primární podoby také dvě alternativní modifikace, které zohledňovaly regionální příslušnost jednotlivých osob. Cílem těchto modifikací bylo zachytit osoby, na které potenciální regionální proměnné působí dlouhodobě. Na základě údajů ohledně změny bydliště osoby od jejího narození a lokalizace bydliště osoby jeden rok před sčítáním tak byly součástí jednotlivých alternativ pouze osoby, u nichž se tyto údaje shodovaly na úrovni krajů či obcí (dále označovány jako modifikace základního souboru „kraj“ a „obec“). Příslušnost jednotlivých osob do konkrétních krajů bylo potřeba na základě odpovídajícího kódu obce nadstavbově stanovit za pomoci číselníku Registru sčítacích obvodů a budov (RSO). S jeho pomocí byly také jednotlivým obcím přiřazeny informace z definitivních výsledků SLDB 2011 týkající se počtu obvykle bydlících v konkrétních obcích, které byly v průběhu analýz využity při zařazování obcí do velikostních kategorií. Propojování základního souboru a číselníků RSO probíhalo stejně jako celý analytický proces v prostředí statistického softwaru SPSS, který umožnil tyto postupy automatizovat. Ručně bylo nutné korigovat pouze nesrovnalosti plynoucí z časového nesouladu mezi daty tvořícími základní soubor platnými k 26. 3. 2011 a aktuálnějšími informacemi z RSO, jehož číselníky reflektují situaci k 1. 1. 2018. Během tohoto více než šestiletého období došlo k zániku obce Brdy, která tak v aktuálních verzích číselníků RSO nefigurovala, a údaje o její příslušnosti do kraje a počtu obvykle bydlících osob byly doplněny manuálně.

Další částí přípravy datového souboru bylo vytvoření reprezentativních výběrových souborů z jednotlivých souborů základních. Analýza výběrových souborů je v obecné rovině běžnou statistickou praxí, pro jejíž provedení nebyla existence samotných základních souborů v našem případě překážkou. Byla naopak bonusem, díky kterému bylo možné výběrové výsledky porovnat s výsledky za základní soubory, a přesvědčit se tak o vhodnosti a správnosti vytvořených výběrů. Výběrové soubory byly vytvořeny tři, analogicky k souboru základnímu a jeho modifikacím „kraj“ a „obec“, přičemž velikost každého z těchto souborů byla 10 000 osob (zvolená hodnota bude blíže diskutována později v souvislosti s víceúrovňovými výběry). Základní výběrový soubor a výběrový soubor „kraj“ byly vytvořeny metodou prostého náhodného výběru. U výběrového souboru „obec“ byl použit stratifikovaný náhodný výběr podle věku jednotlivých osob a velikosti obce, díky kterému jsme byli schopni zohlednit podstatnou skutečnost, že mladší osoby a osoby žijící v populačně velké obci žijí s větší pravděpodobností po celý svůj život v jedné obci než starší osoby a osoby z populačně menších obcí. Neboli jinými slovy že

pravděpodobnost dlouhodobého pobytu na území jedné obce klesá jednak s rostoucím věkem konkrétních osob žijících v této obci, a také se snižující se populační velikostí dané obce.

Vytvoření těchto reprezentativních výběrových souborů bylo předstupněm hlavní a nejdůležitější fáze celé přípravy datového souboru, kterou byla tvorba vícestupňových výběrových souborů. Ty jsou, jak již bylo blíže diskutováno v rámci teoretických kapitol, díky hierarchické struktuře svých dat ideálním podkladem pro víceúrovňové modely. Samotné vytvoření tří typů logicky odpovídajících souborů, které by dávaly v celém kontextu smysl, bylo náročným úkolem, jehož řešení vyžadovalo konzultaci s praxí.

Zajímavé informace a cenná doporučení byly v této věci získány od Ing., Mgr. Martina Zeleného, Ph.D., ředitele sekce demografie a sociálních statistik Českého statistického úřadu (ČSÚ), který poskytl bližší seznámení s metodikou jednoho z nejvýznamnějších šetření u domácností, jež ČSÚ zajišťuje, kterým je šetření Životní podmínky. Jedná se o národní modifikaci celoevropského šetření *European Union – Statistics on Income and Living Conditions* (EU–SILC), které je zaměřené na zjišťování ekonomické a sociální situace domácností ve 34 převážně evropských zemích. V jeho rámci je v Česku každý rok osloveno zhruba 10 000 bytů a jejich obyvatel (průměrná hodnota tzv. *actual sample size* je za poslední tři dostupné roky 10 109 bytů). Tito oslovení však nejsou ze zákona povinni údaje v rámci šetření Životní podmínky poskytovat, jelikož účast v něm je, na rozdíl například od sčítání, dobrovolná. Řada z nich tak vyšetřena není a výsledný výběrový soubor, který je předmětem dalšího studia, tak bez těchto tzv. *nonresponse* tvoří obvykle zhruba pouze kolem 8 000 bytů čítajících 14 000 osob (průměrná hodnota tzv. *achieved sample size* je za analogické roky 2013–2015 8 081 bytů a 14 130 osob) (CIRCABS 2016). Tyto hodnoty tak byly hlavní inspirací při volbě velikostí jak reprezentativních výběrových souborů, jejichž velikost tedy byla 10 000 osob, tak také výběrů vícestupňových. Při jejich tvorbě bylo bráno v potaz také pravidlo týkající se velikosti souboru dat analyzovaných metodami víceúrovňového modelování, jež bylo blíže představeno a diskutováno v rámci teoretických kapitol. Dle tohoto pravidla bychom pro analýzu například 30 nadřazených skupin měli mít k dispozici 30 jedinců v každé z nich, čili dohromady 900 jednotek. Pro analýzu prováděnou v rámci této práce tak byly v kontextu tohoto pravidla a velikosti výběrového souboru šetření EU–SILC hodnoty pro počet nadřazených skupin a jedinců v každé z nich stanoveny na 120. Maximální velikost souboru tak mohla být 14 400 osob.

Ve skutečnosti však byly jednotlivé vícestupňové výběrové soubory tvořeny 7 563–11 094 osobami.

Tato variabilita počtu osob v jednotlivých souborech a zejména také nedosažení maximální možné velikosti výběrových souborů jsou způsobeny použitou metodou výběru. Její základ, jak už samotné označení vícestupňových výběrových souborů napovídá, spočívá v postupném výběru v rámci více stupňů. Takovéto stupně byly v našem případě dva, přičemž na prvním stupni se výběr týkal obcí a na druhém poté jedinců žijících v těchto vybraných obcích. S pomocí prostého náhodného výběru 120 obcí na prvním stupni a následného prostého náhodného výběru 120 jedinců v každé vybrané obci by však výsledný výběrový soubor nerespektoval odlišnou velikost jednotlivých obcí, která má na riziko nezaměstnanosti na individuální úrovni nezpochybnitelný vliv (jak bude blíže diskutováno později v souvislosti s analýzou rizika nezaměstnanosti na individuální úrovni). Z tohoto důvodu bylo potřeba přistoupit k použití tzv. *self-weighting sample*, který odlišnou velikost jednotlivých obcí respektuje skrze rozdílné hodnoty pravděpodobnosti konkrétních obcí být vybrána na prvním stupni (Tambay, Catlin 1995). Tyto pravděpodobnosti přitom v určité míře odrážejí populační velikost jednotlivých obcí v podobě počtu osob žijících v dané obci, které v průběhu předchozích částí přípravy datového souboru prošly všemi prováděnými redukcemi a jsou součástí základního souboru či analogicky jeho modifikací. Prakticky tak byl vícestupňový výběr proveden tak, že ze základních souborů bylo prostým náhodným výběrem vybráno 120 osob. Tímto krokem tak byly přeneseně zohledněny kýžené pravděpodobnosti, jelikož základní soubory byly tvořeny individuálními daty za jednotlivé osoby, kterých bylo z velkých obcí více a z malých úměrně méně. Vlivem této skutečnosti tak také nikdy nebylo vybráno 120 osob žijících ve 120 různých obcích. Dosažené počty se pohybovaly mezi 76 a 99 vybranými obcemi. V každé z takto vybraných obcí pak již byl prostým náhodným výběrem vybrán konstantní počet 120 osob, které tvořili jednotlivé vícestupňové výběrové soubory. V případě, že k vybrané obci příslušelo méně než 121 osob, byli do výběrového souboru zařazeni všichni jedinci.

Jak bylo poznamenáno výše, vytvořeny byly tři typy vícestupňových výběrových souborů, které odpovídaly základnímu souboru a jeho modifikacím „kraj“ a „obec“. Postup výběru ze základního souboru byl zopakován 100krát a bylo tak vytvořeno 100 základních vícestupňových výběrových souborů, jejichž základní charakteristiky jsou k dispozici

v příloze 1. Na základě analýzy variability dále zkoumaného klíčového znaku ICC, jejíž výstup je v podobě histogramu schematicky znázorněn na obrázku 4, bylo v případě souborů „kraj“ a „obec“ víceúrovňových výběrových souborů vytvořeno již pouze deset.

## 5.2 Analýza rizika nezaměstnanosti na individuální úrovni

Předmětem analýzy rizika nezaměstnanosti na individuální úrovni, prováděné tedy v prostředí statistického softwaru SPSS, byl jednak základní soubor a jeho modifikace „kraj“ a „obec“ a dále také reprezentativní výběrové soubory. Postupně byly v rámci jednotlivých souborů studovány vztahy mezi závisle proměnnou nezaměstnanost a výše jmenovanými devíti nezávisle proměnnými. Těmi tedy byly věk, rodinný stav/registrované partnerství, život s partnerem, rodinný vztah, nejvyšší ukončené vzdělání, obor vzdělání, pohlaví, život v domácnosti s alespoň jedním dítětem a náboženská víra osoby. Základní vybrané údaje o jednotlivých proměnných jsou znázorněny v tabulce 2 a podrobnější informace jsou k dispozici v příloze 2.

Kromě základních frekvenčních analýz, jejichž některé výstupy jsou spolu s dalšími výsledky k dispozici v tabulce 4, došlo také na studium vztahů mezi jednotlivými nezávisle proměnnými, v důsledku čehož byl počet nezávisle proměnných snížen na šest. Mezi třemi vyřazenými proměnnými figurovaly jednak proměnné RODSTAV a RODVZTAH a dále také proměnná VZDEL\_OB.

Proměnné RODSTAV a RODVZTAH vykazovaly ještě s proměnnou PARTNER relativně vysoké hodnoty vzájemných závislostí. O vyřazení právě proměnných RODSTAV a RODVZTAH bylo rozhodnuto na základě analýzy závislostí s proměnnou VEKKAT, jejichž hodnoty byly pro RODSTAV a RODVZTAH vyšší než v případě proměnné PARTNER. Přesné hodnoty těchto závislostí jsou prostřednictvím kontingenčních koeficientů vyjádřeny v tabulce 3, která je k dispozici níže.

Také příčinou vyřazení proměnné VZDEL\_OB byla vysoká korelace s jinou proměnnou. Tentokrát se jednalo o proměnnou VZDEL\_ST (kontingenční koeficient = 0,736), která je v porovnání s proměnnou VZDEL\_OB daleko lépe interpretovatelná.



**Tabulka 2: Seznam analyzovaných proměnných**

Proměnná	Název proměnné	Název kategorie
Nezaměstnanost	NEZAM	Zaměstnaní Nezaměstnaní
Věk	VEKKAT	15-24 let 25-34 let 35-44 let 45-54 let 55-64 let
Rodinný stav/registrované partnerství	RODSTAV	Svobodní Ženatí/vdané Rozvedení a ovdovělí
Život s partnerem	PARTNER	Ne Ano
Rodinný vztah	RODVZTAH	Svobodní žijící sami Svobodní žijící v nesezdaném svazku Ženatí/vdané Rozvedení a ovdovělí žijící sami Rozvedení a ovdovělí žijící v nesezdaném svazku
Nejvyšší ukončené vzdělání	VZDEL_ST	ZŠ SŠb SŠm VŠ
Obor vzdělání	VZDEL_OB	Bez vzdělání či základní vzdělání Společenské vědy, nauky a služby Ekonomie Právo Přírodní vědy a nauky Technické vědy a nauky Zemědělsko-lesnické a veterinární vědy a nauky Zdravotnictví, lékařské a farmaceutické vědy a nauky Vojenské vědy a nauky
Pohlaví	POHL	Muž Žena
Život v domácnosti s alespoň jedním dítětem	DETI	Domácnost bez dětí Domácnost s dětmi
Náboženská víra	VIRA	Nevěřící Katolíci Ostatní křesťané Věřící - ostatní náboženství Věřící - nehlásící se k žádné církvi ani nábož. společnosti

**Zdroj:** vlastní zpracování

**Tabulka 3: Kontingenční koeficienty (Pearson Chi-Square) vyjadřující závislost mezi vybranými proměnnými**

	RODSTAV	RODVZTAH	PARTNER	VEKKAT
RODSTAV	x	0,816	0,651	0,535
RODVZTAH	0,816	x	0,707	0,544
PARTNER	0,651	0,707	x	0,365
VEKKAT	0,535	0,544	0,365	x

**Zdroj:** vlastní výpočty

Do pokročilejších analýz rizika nezaměstnanosti na individuální úrovni tak vstupovalo šest nezávislých proměnných. Provedeny byly binární logistické regresní analýzy, jejichž smyslem bylo podrobněji mapovat tvar závislosti mezi těmito proměnnými a vysvětlovanou proměnnou nezaměstnanost. Pro každou jednotlivou nezávisle proměnnou byla nejprve provedena samostatná párová analýza vlivu na závisle proměnnou. Po těchto analýzách přišel na řadu i samotný souhrnný model binární logistické regrese zahrnující všechny vysvětlující proměnné.

Výstupy párových analýz jsou v podobě hodnot poměru šancí také schematicky znázorněny v tabulce 4, a to souhrnně za všechny proměnné. Z těchto hodnot je jasné vidět, u kterých skupin osob panuje vyšší šance na to být nezaměstnaný. Jedná se o mladé osoby ve věkové kategorii 15–24 let, o osoby žijící bez partnera a mající pouze základní vzdělání. Dále jsou vyšším rizikem nezaměstnanosti ohroženy také ženy, lidé žijící v domácnosti bez dětí a věřící osoby nekřesťanského vyznání.

Souhrnný model binární logistické regrese, jehož výhodou oproti předchozím párovým analýzám je zjištění vztahu nezávisle proměnné k riziku nezaměstnanosti při kontrole hodnot všech ostatních nezávislých proměnných, byl konstruován prostřednictvím více různých metod. Pomocí metody Forward Conditional byly identifikovány čtyři nejvýznamnější proměnné, jež mají na individuální úrovni vliv na riziko nezaměstnanosti. Bližší specifikace tohoto modelu zahrnujícího postupně proměnné nejvyšší ukončené vzdělání, život s partnerem, věk a pohlaví, je k dispozici v tabulce 5. Nejvyšší ukončené vzdělání má mezi zahrnutými vysvětlujícími proměnnými klíčové postavení, čemuž odpovídají i rozdíly v hodnotách poměru šancí pro jednotlivé stupně vzdělání, které jsou opravdu markantní. Například pro osoby s vysokoškolským vzděláním je šance na to být nezaměstnaným přibližně desetkrát nižší než u osob spadajících do kategorie ZŠ. Nezanedbatelným faktorem ovlivňujícím riziko nezaměstnanosti je samozřejmě také věk.

U něj můžeme sledovat nejprve snižující se šanci směrem od osob nejmladší věkové kategorie k osobám ve věku 35–44 let a poté naopak vzrůst hodnot dále k vyšším věkovým skupinám. Také to, zda žije daná osoba s partnerem, má na míru rizika nezaměstnanosti vliv. Hodnota poměru šancí je pro osoby žijící s partnerem pouze poloviční v porovnání s osobami žijícími bez partnera. Čtvrtým klíčovým determinantem nezaměstnanosti je dle modelu binární logistické regrese pohlaví, v jehož případě bylo vyšší riziko nezaměstnanosti identifikováno u žen.

**Tabulka 4: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti na individuální úrovni; základní soubor; 2 232 757 osob**

Individuální proměnné		Riziko nezaměstnanosti (základní soubor)		
		%	Poměr šancí	p-hodnota
VEKKAT	15-24 let	25,0	<b>1</b>	0,000
	25-34 let	9,3	<b>0,306</b>	0,000
	35-44 let	6,9	<b>0,221</b>	0,000
	45-54 let	7,7	<b>0,251</b>	0,000
	55-64 let	9,8	<b>0,328</b>	0,000
PARTNER	Ne	14,8	<b>1</b>	0,000
	Ano	6,3	<b>0,387</b>	0,000
VZDEL_ST	ZŠ	31,8	<b>1</b>	0,000
	SŠb	11,4	<b>0,277</b>	0,000
	SŠm	7,4	<b>0,171</b>	0,000
	VŠ	3,6	<b>0,081</b>	0,000
POHL	Muž	8,8	<b>1</b>	0,000
	Žena	10,2	<b>1,173</b>	0,000
DETI	Domácnost bez dětí	10,8	<b>1</b>	0,000
	Domácnost s dětmi	7,4	<b>0,662</b>	0,000
VIRA	Nevěřící	9,4	<b>1</b>	0,000
	Katolíci	8,8	<b>0,931</b>	0,000
	Ostatní křesťané	9,1	<b>0,971</b>	0,064
	Věřící - ostatní náboženství	12,6	<b>1,393</b>	0,000
	Věřící - nehlásící se k žádné církvi ani nábož. společnosti	10,3	<b>1,103</b>	0,000

**Zdroj:** vlastní výpočty

Obdobný postup zahrnující provedení jak základních frekvenčních analýz, tak také samostatných párových analýz a konstrukci souhrnného modelu binární logistické regrese byl aplikován i na dvě alternativní modifikace základního souboru „kraj“ a „obec“. Výstupy analýz těchto souborů jsou k dispozici v přílohách 3 a 4. Výsledky se však v žádném ohledu významně neodlišují od výsledků analýz základního souboru.

Jak bylo řečeno výše, předmětem analýzy rizika nezaměstnanosti na individuální úrovni byly kromě základního souboru a jeho modifikací „kraj“ a „obec“ také reprezentativní výběrové soubory. Výstupy analýz těchto souborů jsou pro přehlednost k dispozici pouze v přílohách 5, 6 a 7, kde jsou k porovnání znovu znázorněny i tabulky za základní soubory. Z tohoto porovnání je patrné, že výsledné hodnoty analyzovaných ukazatelů nevykazují napříč oběma typy souborů velké rozdíly. Pouze v případě reprezentativního výběrového souboru „obec“ se vztah mezi nezaměstnaností a pohlavím jevil opačně než u všech ostatních souborů. U žen tak byla vykazovaná šance na to být osobou nezaměstnanou nižší než u mužů a to o 2,8 %. Tato skutečnost tak dokládá nejslabší vztah proměnné pohlaví k riziku nezaměstnanosti ze všech sledovaných proměnných. V celkovém pohledu však výsledky analýz reprezentativních výběrových souborů potvrdily výsledky získané na úrovni souborů základních, jak bylo očekáváno.

**Tabulka 5: Model binární logistické regrese; základní soubor; 2 232 757 osob**

Odhady parametrů		Riziko nezaměstnanosti (základní soubor)				
		B	Poměr šancí	95% interval spolehlivosti pro poměr šancí		p-hodnota
Absolutní člen		0,074	1,077			0,000
VZDEL_ST	ZŠ		1			
	SŠb	-1,088	0,337	0,332	0,342	0,000
	SŠm	-1,689	0,185	0,182	0,187	0,000
	VŠ	-2,295	0,101	0,099	0,103	0,000
VEKKAT	15-24 let		1			
	25-34 let	-0,724	0,485	0,478	0,492	0,000
	35-44 let	-0,976	0,377	0,371	0,383	0,000
	45-54 let	-0,906	0,404	0,398	0,411	0,000
	55-64 let	-0,653	0,520	0,511	0,530	0,000
PARTNER	Ne		1			
	Ano	-0,697	0,498	0,493	0,503	0,000
POHL	Muž		1			
	Žena	0,291	1,338	1,325	1,351	0,000

**Zdroj:** vlastní výpočty

Veškeré analýzy rizika nezaměstnanosti na individuální úrovni tak prokázaly, že determinanty mající na nezaměstnanost vliv skutečně na této úrovni existují. Pomocí modelu binární logistické regrese byly identifikovány čtyři nejvýznamnější proměnné hrající roli těchto determinantů. Výsadní postavení má mezi nimi nejvyšší ukončené vzdělání, které je proměnnou mající na riziko nezaměstnanosti na individuální úrovni enormní vliv. Významný však je i vliv dalších proměnných, zejména života s partnerem a věku.

### 5.3 Odhalení regionálních vlivů na riziko nezaměstnanosti

Předmětem analýzy směřující k odhalení regionálních vlivů na riziko nezaměstnanosti byly v různých fázích tohoto procesu veškeré dostupné soubory. Nejprve byly za tímto účelem podrobněji studovány soubory základní a reprezentativní výběrové. Ty byly podrobeny obdobným frekvenčním a párovým analýzám, které byly popsány v předchozí části, avšak přirozeně s jinými vysvětlujícími proměnnými zohledňujícími právě regionální rozměr. Poté již přišla na řadu i práce s vícestupňovými výběrovými soubory. Klíčovou nezávisle proměnnou frekvenčních a párových analýz byla jednak příslušnost jednotlivých osob do konkrétních krajů (KRAJ) a dále také velikostní kategorie obcí (VELKAT).

V případě proměnné KRAJ rozlišovaly analýzy mezi všemi existujícími kraji bez výjimky, a proměnná tak pracovala se 14 kategoriemi. Výsledky těchto analýz potvrdily teoretické poznatky z oblasti studia nezaměstnanosti. Nejnižší riziko nezaměstnanosti, vyjádřené hodnotou poměru šancí, mají ve většině případů osoby žijící v Praze, ve Středočeském či v Plzeňském kraji. Na opačné straně spektra se pak nejčastěji nachází kraj Ústecký, Moravskoslezský, Karlovarský či Olomoucký, kde výstupy párových analýz obvykle ukazují o minimálně 50 % vyšší šanci na to být nezaměstnanou osobou než v Praze. Konkrétní hodnoty jsou k dispozici v tabulce 6 a přílohách 8, 9, 10 a 11.

**Tabulka 6: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující krajskou příslušnost; základní soubor; 2 232 757 osob**

Individuální proměnné – regionální vliv		Riziko nezaměstnanosti		
		%	Poměr šancí	p-hodnota
KRAJ	Hlavní město Praha	6,6	<b>1</b>	0,000
	Jihočeský kraj	8,4	<b>1,287</b>	0,000
	Jihomoravský kraj	10,2	<b>1,599</b>	0,000
	Karlovarský kraj	11,3	<b>1,797</b>	0,000
	Kraj Vysočina	9,1	<b>1,419</b>	0,000
	Královéhradecký kraj	8,1	<b>1,244</b>	0,000
	Liberecký kraj	10,5	<b>1,657</b>	0,000
	Moravskoslezský kraj	11,6	<b>1,845</b>	0,000
	Olomoucký kraj	11,3	<b>1,797</b>	0,000
	Pardubický kraj	8,7	<b>1,347</b>	0,000
	Plzeňský kraj	8,0	<b>1,232</b>	0,000
	Středočeský kraj	7,9	<b>1,215</b>	0,000
	Ústecký kraj	12,6	<b>2,027</b>	0,000
	Zlínský kraj	9,7	<b>1,517</b>	0,000

**Zdroj:** vlastní výpočty

Ordinální proměnná VELKAT byla uspořádána do pěti kategorií. Do první kategorie spadaly osoby žijící v obcích s méně než 500 obyvateli, další kategorie byly vymezeny s hranicemi 500–1 999 obyvatel, 2 000–9 999 obyvatel, 10 000–99 999 obyvatel a více než 99 999 obyvatel. Právě v největších obcích se statistickými počty obyvatel je, dle výstupů schematicky zobrazených v tabulce 7 a přílohách 12, 13, 14 a 15, situace týkající se nezaměstnanosti nejpriznivější. Rozdíly v hodnotách poměru šancí jsou však znatelně menší než v případě krajské diferenciaci. Nejodlehlejší byla hodnota 0,687 ukazující na 1,5krát nižší šanci na to být nezaměstnanou osobou v takovémto velkoměstě než v malé obci s méně než 500 obyvateli. Tyto malé obce však obvykle nevykazovaly z hlediska rizika nezaměstnanosti nejhorší výsledky, jak by se možná mohlo na první pohled zdát. V pěti ze šesti případů byly nejvyšší hodnoty poměru šancí pozorovány u středně velkých obcí s 10 000–99 999 obyvateli. Tyto pozorované rozdíly se však většinou pohybovaly pouze v řádu jednotek procent a nebyly tudíž příliš významné.

**Tabulka 7: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující velikostní kategorie obcí; základní soubor; 2 232 757 osob**

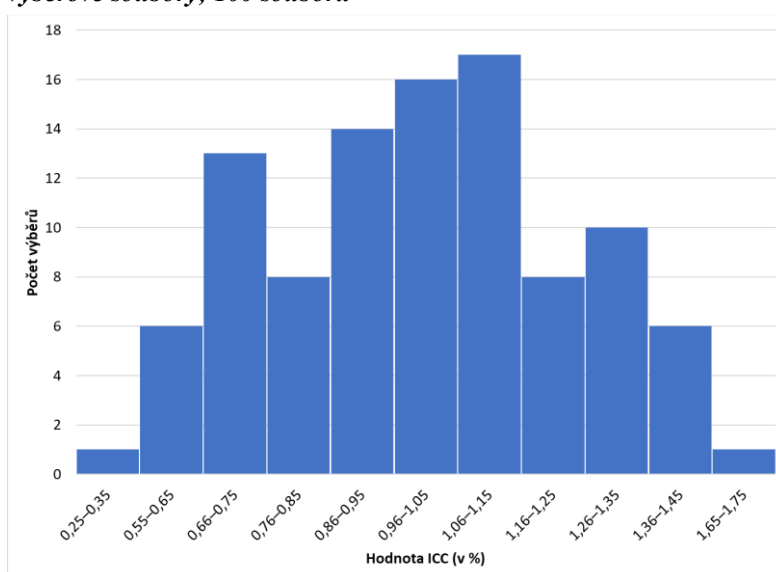
Individuální proměnné – regionální vliv		Riziko nezaměstnanosti		
		%	Poměr šancí	p-hodnota
VELKAT	méně než 500 obyvatel	10,3	<b>1</b>	0,000
	500–1 999 obyvatel	9,5	<b>0,915</b>	0,000
	2 000–9 999 obyvatel	9,4	<b>0,909</b>	0,000
	10 000–99 999 obyvatel	10,3	<b>1,001</b>	0,918
	více než 99 999 obyvatel	8,0	<b>0,758</b>	0,000

**Zdroj:** vlastní výpočty

Náplní práce s víceúrovňovými výběrovými soubory již byly pokročilejší analýzy. S cílem odhalit regionální vlivy na riziko nezaměstnanosti a zejména také rozhodnout, zda má opravdu smysl přistoupit ke konstrukci víceúrovňového modelu, byla pro veškeré víceúrovňové výběrové soubory spočítána hodnota ICC. Kritickou hranicí je zde, jak již víme z diskuze v rámci teoretických kapitol, hodnota 5 %, jejíž dosažení je signálem pro opodstatněné použití víceúrovňových modelů. V případě souborů studovaných v rámci této práce však této hodnoty dosaženo nebylo. Na rozdíly mezi obcemi, které byly v průběhu této analýzy chápány jako nadřazené skupiny, připadalo u víceúrovňových výběrových souborů odpovídajících základnímu souboru v průměru pouze 1 % z celkového rozptylu. Soubory odpovídající modifikaci „kraj“ vykazovaly průměrnou hodnotu 0,75 % a v případě souborů odpovídajících modifikaci „obec“ bylo dosaženo hodnoty 0,76 %. Histogramy znázorňující rozložení všech hodnot jsou znázorněny na

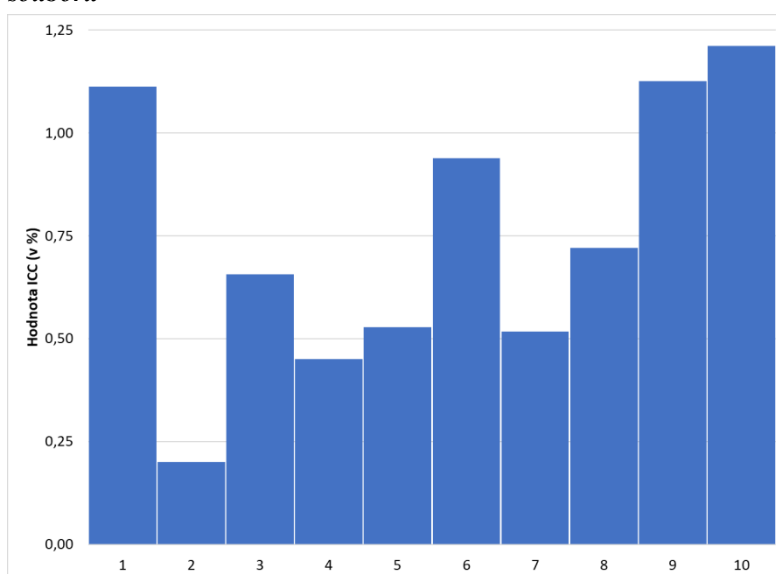
obrázcích 4, 5 a 6. Argument pro opodstatněné použití víceúrovňových modelů byl kromě teoretické roviny reprezentované kritickou hranicí 5 % diskutován také v rovině empirické. V rámci již uskutečněných studií aplikujících víceúrovňově konstruované modely, které byly blíže představeny dříve, bylo pátráno po hodnotách ICC. Ani tyto hodnoty však nebyly pádným důvodem aplikace víceúrovňově konstruovaného modelu v případě souborů studovaných v rámci této práce, jelikož skutečně přesahovaly teoretickou hranici 5 %. Analýzy Vody (2013) či Rychtaříkové (2015) oscilovaly kolem 10 %, Soukup (2006) dosáhl ve své studii dokonce 33 %.

**Obrázek 4:** Histogram znázorňující rozložení hodnot ICC; základní vícestupňové výběrové soubory; 100 souborů



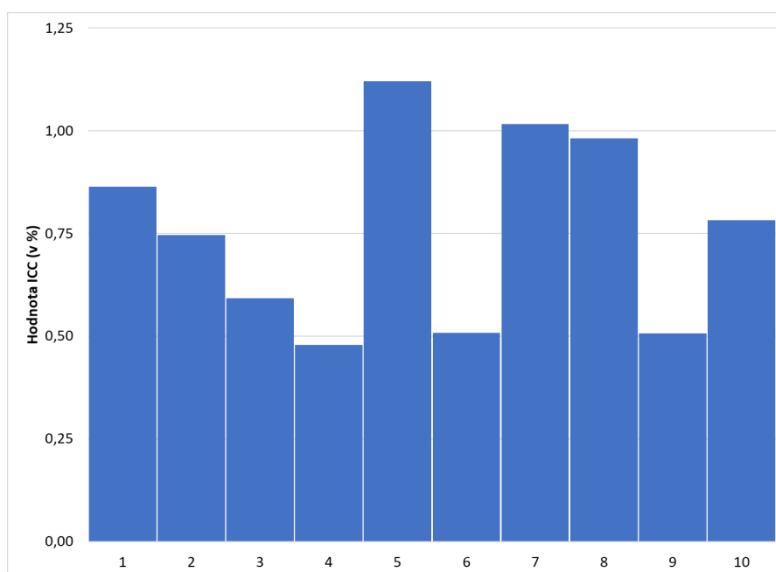
**Zdroj:** vlastní výpočty

**Obrázek 5:** Graf znázorňující hodnoty ICC; vícestupňové výběrové soubory „kraj“; 10 souborů



**Zdroj:** vlastní výpočty

**Obrázek 6:** Graf znázorňující hodnoty ICC; vícestupňové výběrové soubory „obec“; 10 souborů



**Zdroj:** vlastní výpočty

Veškeré analýzy směřující k odhalení regionálních vlivů na riziko nezaměstnanosti tak ukázaly, že determinanty nezaměstnanosti mající regionální povahu skutečně existují. Jejich vliv na riziko nezaměstnanosti je však statisticky významně nižší než vliv individuálních proměnných. Určitý větší význam můžeme přikládat příslušnosti jednotlivých osob do konkrétních krajů, zohledněné vysvětlující proměnnou KRAJ. V případě proměnné VELKAT je vliv opravdu již skoro zanedbatelný, jelikož přirozené rozdíly mezi obcemi různých velikostních kategorií jsou často mazány dojížděnou za prací mimo obec svého obvyklého bydliště.

## 5.4 Víceúrovňová analýza rizika nezaměstnanosti

Pokročilejší víceúrovňově konstruovaná analýza rizika nezaměstnanosti nemá v případě našich datových souborů, jak již bylo naznačeno v předchozí části, smysl. Nulové modely sestavené pro vícestupňové výběrové soubory nepoukázaly na dostatečný vliv hierarchické struktury dat, který by byl kýženým opodstatněním pro aplikaci víceúrovňové analýzy. O vhodnosti jejího použití při studiu rizika nezaměstnanosti jsme však i nadále přesvědčeni. Napovídají mu nejen výše zmiňované úspěšné aplikační příklady, ale i výstupy provedených analýz, které na určitý regionální vliv poukázaly. Jeho nedostatečné, statisticky nevýznamné projevení tak sice bylo na jedné straně důvodem pro nepoužití víceúrovňové analýzy studovaných souborů, zároveň však ale bylo počátkem dalšího bádání. V jeho průběhu byla ještě nad rámec původně zamýšleného obsahu práce



vytvořena obdoba základního souboru zahrnující však pouze osoby s vysokoškolským vzděláním. Tímto krokem tak došlo k eliminaci vlivu nejvýznamnějšího individuálního determinantu nezaměstnanosti, tedy nejvyššího ukončeného vzdělání. Stejným způsobem, jakým byly ze základního souboru a jeho modifikací vytvořeny soubory víceúrovňové výběry, došlo i ke vzniku deseti víceúrovňových výběrových souborů vysokoškoláků. Tyto soubory však nebyly dále podrobovány kompletnímu výše popisovanému analytickému procesu, ale staly se pouze předmětem analýzy ICC, která by mohla odhalit významnější regionální vlivy v takto zúženém souboru. Ani v tomto případě však bohužel nebylo kýmžých 5 % dosaženo, a do budoucna tak nezbývá nic jiného, než se pokoušet hledat další nové cesty vedoucí ke smysluplné víceúrovňové analýze rizika nezaměstnanosti a podchycení regionálních vlivů, které ho ovlivňují.

## Kapitola 6

### Závěr

Podrobně představit statistickou metodu víceúrovňového modelování, která je v geografickém výzkumu používána zatím velice okrajově, bylo hlavním cílem této bakalářské práce. Jeho naplnění tak snad představuje drobný střípek do mozaiky vědeckých, metodologicky orientovaných prací, kterým je mezi odbornou, geografickou veřejností věnována menší pozornost než pracím empiricky zaměřeným. Nejednalo se však pouze o teoretické představení uvedené metody, ale zejména o diskuzi jejího aplikačního potenciálu včetně ukázky její aplikace v konkrétním geografickém výzkumu, která je podle mého názoru nedílnou součástí komplexního pojetí jakéhokoli metodologického tématu. Zároveň je snahou o vyváženější a rovnocennější přístup k metodologicky a empiricky zaměřeným geografickým výzkumným pracím, které mají obě své nezastupitelné místo a navzájem by se neměly z centra geografické pozornosti vytlačovat, nýbrž však se doplňovat, jelikož jedna bez druhé nemá smysl natož pak šanci na úspěch.

Kýžená komplexnost kvalitního pojetí problematiky víceúrovňového modelování zahrnuje celou řadu dílčích oblastí, které by v jeho rámci neměly být vynechány. Více než žádoucí je úvodní ukotvení konkrétní metody v rámci širší metodologické typologie, které poskytne první bližší vymezení a seznámení s vnějšími okolnostmi. Víceúrovňové modelování je v tomto smyslu pevně zakotveno mezi kvantitativními metodami, které procházejí v sociálně-geografickém výzkumu značně progresivním vývojem. V jeho průběhu je však víceúrovňové modelování jako neprostorová metoda spíše upozaděno, jelikož právě prostorové aspekty geografických dat jsou tím, co stále silněji přitahuje pozornost mnoha geografů. Výsledkem je pak vznik celé řady nových metod vyvinutých speciálně pro geografické výzkumy právě s ohledem na ony prostorové aspekty dat (například geograficky vážená regrese). Přestože víceúrovňové modelování tedy není ve své podstatě

metodou prostorovou, může ryze prostorovým metodám konkurovat, neboť je v rámci svých více úrovní schopno prostorovost a regionalitu reflektovat.

Podrobnější diskuze podstaty a jednotlivých motivací použití metody víceúrovňového modelování je již další oblastí zasluhující si v rámci komplexního metodologického pojetí svůj prostor. Určitým vyvrcholením této teoretické části je poté popis samotné konstrukce víceúrovňových modelů. Po něm přicházejí na řadu oblasti zabývající se praxí a aplikační stránkou celé problematiky. Ukázky jak již uskutečněných studií aplikujících víceúrovňově konstruované modely, tak také vlastní analýza jsou stejně významnými oblastmi komplexního pojetí představení metody víceúrovňového modelování jako oblasti předchozí. Zejména ve vlastní analýze je pak také příležitost v praxi otestovat možnosti a omezení představované metody.

Analýza rizika nezaměstnanosti přinesla několik zajímavých poznatků, a splnila tak další dílčí cíl této bakalářské práce, kterým byla identifikace jednotlivých individuálních a regionálních proměnných souvisejících s nezaměstnaností a odhalení velikosti efektu těchto proměnných na nezaměstnanost. Nejzásadnější vliv byl mezi individuálními proměnnými prokázán u nejvyššího ukončeného vzdělání, věku, dále u proměnné život s partnerem a také u pohlaví. Velikost efektu těchto individuálních proměnných byla přitom výrazně větší než v případě proměnných kontextuálních. Mezi těmi stojí za zmínku pouze regionální proměnná KRAJ zohledňující příslušnost jednotlivých osob do konkrétních krajů.

Pokročilejší analýza související s vlivem regionálních determinantů na riziko nezaměstnanosti bohužel neměla v případě námi analyzovaných souborů, ve světle celkového zaměření této práce, příliš pozitivní výsledek. Vykazovala totiž hodnoty hovořící pro nevhodnost použití víceúrovňově konstruovaných modelů. Příčinou tohoto „popření“ hierarchické struktury dat, o níž jsme na základě literatury a analýz na individuální úrovni přesto přesvědčeni, je nespíš způsob tvorby jednotlivých souborů, jejichž reprezentativnost nemůže být bohužel zaručena. Svoji roli dále jistě může hrát také způsob vymezení regionálního kontextu, tedy v našem případě krajů. Zúžení sledovaného souboru osob pouze na vysokoškolsky vzdělané osoby na statisticky významnější vliv regionálních faktorů neukázalo, nicméně analýzy dalších subpopulací z hlediska jejich rizika nezaměstnanosti by jistě byly velmi přínosné a mohly by ukázat na zajímavé

---

souvislosti. Naznačeny zde jsou některé z možností, kterým by bylo záhodno se v dalších výzkumech individuálních a regionálních determinantů rizika nezaměstnanosti věnovat a které by mohly vést k plnému využití potenciálu víceúrovňových modelů.

## SEZNAM LITERATURY

ALKER, H. R. (1969): A Typology of Ecological Fallacies. In: Dogan, M., Rokkan, S. (ed.): Quantitative Ecological Analysis. Massachusetts Institute of Technology, Cambridge, 69–86.

ARNTZ, M., WILKE, R. A. (2009): Unemployment Duration in Germany: Individual and Regional Determinants of Local Job Finding, Migration and Subsidized Employment. *Regional Studies*, 43, 1, 43–61.

ASHBY, W. M. (1965): Soustavy a informace. Vesmír, Praha.

BAUM, S., BILL, A., MITCHELL, W. F. (2009): Employability and Labour Underutilization in Non-Metropolitan Labour Markets. *Regional Studies*, 43, 8, 1091–1103.

BAUM, S., MITCHELL, W. F. (2010): People, Space and Place: a Multidimensional Analysis of Unemployment in Metropolitan Labour Markets. *Geographical Research*, 48, 1, 13–23.

BILLY, J. O. G., MOORE, D. E. (1992): A Multilevel Analysis of Marital and Nonmarital Fertility in the U.S. *Social Forces*, 70, 4, 977–1011.

BOCK, R. D., (1989): Multilevel Analysis of Educational Data. Academic Press, Inc., San Diego.

BRAUN, H. I., JONES, D. H., RUBIN, D. B., THAYER, D. T. (1983): Empirical Bayes Estimation of Coefficients in the General Linear Model with Data of Deficient Rank. *Psychometrika*, 48, 2, 171–182.

BROWN, S., SESSION, J. G. (1997): A Profile of UK Unemployment: Regional versus Demographic Influences. *Regional Studies*, 31, 4, 351–366.

BURIÁNEK, J. (2017): Sociologická encyklopedie.

[https://encyklopedie.soc.cas.cz/w/Anal%C3%BDza\\_kontextu%C3%A1ln%C3%AD](https://encyklopedie.soc.cas.cz/w/Anal%C3%BDza_kontextu%C3%A1ln%C3%AD) (cit. 5. 4. 2018)

CIRCABS (2016): Eurostat, EU–SILC.

[https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp?FormBanner:\\_idcl=navigationTitle&FormBanner\\_SUBMIT=1&org.apache.myfaces.trinidad.faces.STATE=DUMMY&id=853b48e6-a00f-4d22-87db-c40bafd0161d&javax.faces.ViewState=JU5nK7DpKFHnDaNoDO0LCpFqMcxh3aL09eZEUeVjNpux9r4VyJwDDdICshQJun9mxa502cnWX1W4Ka1XR00OTTT0pMz77N3JO1rBVHEe4MZVpWkssxWGvYL4G9ggPGWmbyLmU%2Fn9bxxhEFhhU8FOwuTgMcY%3D](https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp?FormBanner:_idcl=navigationTitle&FormBanner_SUBMIT=1&org.apache.myfaces.trinidad.faces.STATE=DUMMY&id=853b48e6-a00f-4d22-87db-c40bafd0161d&javax.faces.ViewState=JU5nK7DpKFHnDaNoDO0LCpFqMcxh3aL09eZEUeVjNpux9r4VyJwDDdICshQJun9mxa502cnWX1W4Ka1XR00OTTT0pMz77N3JO1rBVHEe4MZVpWkssxWGvYL4G9ggPGWmbyLmU%2Fn9bxxhEFhhU8FOwuTgMcY%3D) (cit. 11. 7. 2018)

CLEVELAND, W. S. (1979): Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American statistical association*, 74, 368, 829–836.

CUBBIN, C., LECLERE, F. B., SMITH, G. S. (2000): Socioeconomic Status and Injury Mortality: Individual and Neighbourhood Determinants. *Journal of Epidemiology & Community Health*, 54, 7, 517–524.

DAVEY SMITH, G., HART, C., WATT, G., HOLE, D., HAWTHORNE, V. (1998): Individual Social Class, Area-based Deprivation, Cardiovascular Disease Risk Factors, and Mortality: The Renfrew and Paisley Study. *Journal of Epidemiology & Community Health*, 52, 6, 399–405.

DIEZ-ROUX, A., NIETO, F., MUNTANER, C., TYROLER, H. A., COMSTOCK, G. W., SHAHAR, E., COOPER, L. S., WATSON, R. L., SZKLO, M. (1997): Neighborhood Environments and Coronary Heart Disease: A Multilevel Analysis. *American Journal of Epidemiology*, 146, 1, 48–63.

DISMAN, M. (2002): Jak se vyrábí sociologická znalost. Univerzita Karlova, Praha.

DRIBE, M., JUÁREZ, S. P., SCALONE, F. (2015): Is It Who You Are or Where You Live? Community Effects on Net Fertility at the Onset of Fertility Decline: A Multilevel Analysis Using Swedish Micro-Census Data. *Population, Space and Place*, 23, 2, 1–27.

DUNCAN, C., JONES, K., MOON, G. (1999): Smoking and Deprivation: Are There Neighbourhood Effects? *Social Science & Medicine*, 48, 4, 497–505.

EZRA, M., KIROS, G.-E. (2006): Rural Out-migration in the Drought Prone Areas of Ethiopia: A Multilevel Analysis. *International Migration Review*, 35, 3, 749–771.

FOTHERINGHAM, A. S., BRUNSDON, C., CHARLTON, M. (2000): *Quantitative Geography: Perspectives on Spatial Data Analysis*. Sage Publications, London.

FOTHERINGHAM, A. S., BRUNSDON, C., CHARLTON, M. (2002): *Geographically Weighted Regression: The Analysis of Spatially Varying Relationship*. John Wiley & Sons, London.

FREEDMAN, D. A. (1999): Ecological Inference and the Ecological Fallacy. In: Smelser, N. J., Baltes, P. B. (ed.): *International Encyclopedia of the Social & Behavioral Sciences* 6: 4027–30. Department of Statistics University of California, Berkeley, 1–6.

GELMAN, A., HILL, J. (2007): *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York.

GLASS, G. V. (1976): Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, 5, 10, 3–8.

GOULD, M. I., FIELDHOUSE, E. (1997): Using the 1991 Census SAR in a multilevel analysis of male unemployment. *Environment and Planning A*, 29, 4, 611–628.

GREGORY, D., JOHNSTON, R., PRATT, G., WATTS, M., WHATMORE, S. (2009): *The Dictionary of Human Geography*. Wiley–Blackwell, Chichester.

HAMPL, M. (1998): *Realita, společnost a geografická organizace: hledání integrálního řádu*. Přírodovědecká fakulta Univerzity Karlovy, Praha.

HAMPLOVÁ, D. (2006): Životní spokojenost, štěstí a rodinný stav v 21 evropských zemích. *Sociologický časopis*, 42, 1, 35–55.

HAMPLOVÁ, D., ŘEHÁKOVÁ, B. (2006): Mimomanželská plodnost: individuální charakteristiky žen a vliv regionu. In: Hamplová, D. (ed): *Mimomanželská plodnost v České republice po roce 1989: sociální a ekonomické souvislosti*. Sociologický ústav Akademie věd České republiky, Praha, 26–39.

HECK, R. H., THOMAS, S. L., TABATA, L. N. (2012): *Multilevel Modeling of Categorical Outcomes Using IBM SPSS*. Routledge, New York.

HENDL, J. (2004): *Přehled statistických metod zpracování dat: analýza a metaanalýza dat*. Portál, Praha.

HONG, I., JUNG, J.-K. (2017): What Is So "Hot" in Heatmap? Qualitative Code Cluster Analysis with Foursquare Venue. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 52, 4, 332–348.

HOX, J. J. (2010): *Multilevel Analysis: Techniques and Applications*. Routledge, New York.

HSU, CH.-Y., CHANG, S.-S., YIP, P. (2016): Individual-, Household- and Neighbourhood-Level Characteristics Associated with Life Satisfaction: A Multilevel Analysis of a Population-based Sample from Hong Kong. *Urban Studies*, 54, 16, 3700–3717.

HUMMELL, H. J., SODEUR, W. (1997): Structural Analysis of Social Networks with Respect to Different Levels of Aggregation. *Mathématiques et sciences humaines*, 35, 137, 37–60.

KAWACHI, I., KENNEDY, B. P., GLASS, R. (1999): Social Capital and Self-Rated Health: A Contextual Analysis. *American Journal of Public Health*, 89, 8, 1187–1195.

KHATTAB, N. (2006): Ethnic and Regional Determinants of Unemployment in the Israeli Labour Market: A Multilevel Model. *Regional Studies*, 40, 1, 93–105.

KIM, D., SUBRAMANIAN, S. V., KAWACHI, I. (2006): Bonding Versus Bridging Social Capital and Their Associations with Self Rated Health: A Multilevel Analysis of 40 US Communities. *Journal of Epidemiology & Community Health*, 60, 2, 116–122.

KOSTELECKÝ, T., ČERMÁK, D. (2003): Výběrová šetření a analýza agregátních dat – diskuse na téma použitelnosti různých přístupů v komparativních analýzách politického chování. *Sociologický časopis*, 39, 4, 529–550.

KOUDELKA, P. (2011): *Demografické aplikace vícestupňových a víceprocesových modelů*.



KLEINSCHMIDT, I., HILLS, M., ELLIOTT, P. (1995): Smoking Behavior Can Be Predicted by Neighborhood Deprivation Measures. *Journal of Epidemiology & Community Health*, 49, 6, 72–77.

KREFT, I., DE LEEUW, J. (1998): *Introducing Multilevel Modeling*. Sage Publications, London.

KUDRNÁČ, A. (2017): Vliv klimatu školní třídy a jejího socioekonomického složení na občanské znalosti a postoj k volební účasti. *Sociologický časopis*, 53, 2, 209–240.

KULU, H., BILLARI, F. (2010): Multilevel Analysis of Internal Migration in a Transitional Country: The Case of Estonia. *Regional Studies*, 38, 6, 679–696.

LAZARSFELD, P. F., MENZEL, H. (1961): On the relation between individual and collective properties. In: Etzioni, A. (ed.): *Complex Organisations: A Sociological Reader*. The Macmillan Company, New York, 422–440.

LEDERMAN, D. (2010): An International Multilevel Analysis of Product Innovation. *Journal of International Business Studies* (2010) 41, 4, 606–619.

LING, B., GUO, Y., CHEN, D. (2018): Change Leadership and Employees' Commitment to Change: A Multilevel Motivation Approach. *Journal of Personnel Psychology*, 17, 2, 85–95.

LUKE, D. A. (2004): *Multilevel Modeling*. Sage University Papers Series on Quantitative Applications in the Social Sciences. Thousand Oaks.

LUTONSKÝ, M. (2011): Co je to Foursquare: návod v češtině.

<https://www.zive.cz/bleskovky/co-je-to-foursquare-navod-v-cestine/sc-4-a-156912/default.aspx> (cit. 13. 4. 2018)

MATTESON, D., BURR, J., MARSHALL, J. (1998): Infant Mortality: A Multi-level Analysis of Individual and Community Risk Factors. *Social Science & Medicine*, 47, 11, 1841–1854.

MOURA, K., SILVEIRA-NETO, R. (2015): Individual and Contextual Determinants of Victimization in Brazilian Urban Centres: A Multilevel Approach. *Urban Studies*, 53, 8, 1559–1573.

NEZLEK, J. B. (2008): An Introduction to Multilevel Modeling for Social and Personality Psychology. *Social and Personality Psychology Compass*, 2, 2, 842–860.

NOVÝ, M. (2015): Politika v supermarketu: Bojkotující spotřebitelé jako agenti sociální změny. *Sociológia*, 47, 1, 87–112.

O'CAMPO, P., XUE, X., WANG, M., CAUGHY, M. (1997): Neighborhood Risk Factors for Low Birthweight in Baltimore: A Multilevel Analysis. *American Journal of Public Health*, 87, 7, 1113–1118.

PETRUSEK, M. (1993): *Teorie a metoda v moderní sociologii*. Karolinum, Praha.

PECHA, O. (2009): *Multilevel Structural Analysis of the Attitudes toward Physical Education: Methodological Consideration*. Dizertační práce. Fakulta tělesné výchovy a sportu UK, Praha.

ROBERTS, E. M. (1997): Neighborhood Social Environments and the Distribution of Low Birth Weight in Chicago. *American Journal of Public Health*, 87, 4, 597–603.

ROBINSON, W. S. (1950): Ecological Correlations and the Behavior of Individuals. *American Sociological Review*, 15, 3, 351–357.

RUBIN, D. B. (1980): Using Empirical Bayes Techniques in the Law School Validity Studies. *Journal of the American Statistical Association*, 75, 372, 801–816.

RUNDLE, A., DIEZ-ROUX, A., FREEMAN, L. M., MILLER, D., NECKERMAN, K. M., WEISS, C. C. (2007): The Urban Built Environment and Obesity in New York City: A Multilevel Analysis. *American Journal of Health Promotion*, 21, 4, 326–334.

RYCHTAŘÍKOVÁ, J. (2015): Délka života, zdraví a postoje v Evropské Unii. *Geografie*, 120, 4, 542–563.

ŘEHÁKOVÁ, B. (1982): Příčinné závislosti a jejich statistický odraz. *Sociologický časopis*, 18, 5, 532–543.

SAMPSON, R. J., RAUDENBUSH, S. W., EARLS, F. (1997): Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy. *Science*, 277, 5328, 918–924.

SCHYNS, B., VAN VELDHOFEN, M. J. P. M. (2010): Group Leadership Climate and Individual Organizational Commitment: A Multilevel Analysis. *Journal of Personnel Psychology* 9, 2, 57–68.

SCHWARTZ, S (1994): The Fallacy of the Ecological Fallacy: The Potential Misuse of a Concept and the Consequences. *American Journal of Public Health*, 84, 5, 819–824.

SNIJDERS, T. A. B., BOSKER, R. J. (2012): Multilevel analysis: an introduction to basic and advanced multilevel modeling. Sage Publications, Los Angeles.

SOUKUP, P. (2006): Proč užívat hierarchické lineární modely? *Sociologický časopis*, 42, 5, 987–1012.

SPILKOVÁ, J., DZÚROVÁ, D., PIKHART, H. (2011): Inequalities in smoking in the Czech Republic: Societal or individual effects? *Health & Place*, 17, 1, 215–221.

SPURNÁ, P. (2006): Současné trendy v kvantitativní analýze geografických dat. Diplomová práce. Katedra sociální geografie a regionálního rozvoje PřF UK, Praha.

STRAKOVÁ, J. (2010): Dopad diferenciací vzdělávacích příležitostí v povinném vzdělávání na vývoj nerovnoměrností ve výsledcích žáků v ČR po roce 2000. *Pedagogika*, 60, 1, 21–37.

SUBRAMANIAN, S. V., KIM, D. J., KAWACHI, I. (2002): Social Trust and Self-Rated Health in US Communities: A Multilevel Analysis. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 79, 4, 21–34.

SUNDQUIST, J., MALMSTROM, M., JOHANSSON, S. E. (1999): Cardiovascular Risk Factors and the Neighbourhood Environment: A Multilevel Analysis. *International Journal of Epidemiology*, 28, 5, 841–845.

TAMBAY, J.-L., CATLIN, G. (1995): Sample Design of the National Population Health Survey. *Health Reports*, 7, 1, 29 – 38.

TIMOFEEVA, A., TESSELKINA, K. (2017): Geographically weighted factor analysis: optimal bandwidth selection. In: *Information Technology and Nanotechnology – 2017, Image Processing and Geoinformation Technology*. Novaja tekhnika, Samara, 663–669.

VODA, P. (2013): Využití víceúrovňového modelování v geografii: Studie spokojenosti se životem v obcích na Broumovsku. In: Robert Osman (ed): Geografický výzkum: společnost a příroda v období krize. Masarykova univerzita, Brno, 236–250.

YANG, X., GUO, F. (1999): Gender Differences in Determinants of Temporary Labor Migration in China: A Multilevel Analysis. *The International Migration Review*, 33, 4, 929–953.

ZHANG, L., MESSNER, S. F., LIU, J. (2007): A Multilevel Analysis of the Risk of Household Burglary in the City of Tianjin, China. *The British Journal of Criminology*, 47, 6, 918–937.

## PŘÍLOHY

**Příloha 1a:** Základní charakteristiky týkající se velikosti základních víceúrovňových výběrových souborů

**Příloha 1b:** Histogram znázorňující rozložení velikostí základních víceúrovňových výběrových souborů; základní víceúrovňové výběrové soubory; 100 souborů

**Příloha 2:** Seznam analyzovaných proměnných včetně podrobnějšího popisu

**Příloha 3:** Výstupy frekvenčních a párových analýz rizika nezaměstnanosti na individuální úrovni; soubor „kraj“ – 1 425 914 osob, soubor „obec“ – 877 064 osob

**Příloha 4:** Model binární logistické regrese; soubor „kraj“ – 1 425 914 osob, soubor „obec“ – 877 064 osob

**Příloha 5:** Výstupy frekvenčních a párových analýz rizika nezaměstnanosti na individuální úrovni; základní soubor – 2 232 757 osob, základní reprezentativní výběrový soubor – 10 000 osob

**Příloha 6:** Výstupy frekvenčních a párových analýz rizika nezaměstnanosti na individuální úrovni; soubor „kraj“ – 1 425 914 osob, reprezentativní výběrový soubor „kraj“ – 10 000 osob

**Příloha 7:** Výstupy frekvenčních a párových analýz rizika nezaměstnanosti na individuální úrovni; soubor „obec“ – 877 064 osob, reprezentativní výběrový soubor „obec“ – 10 000 osob

**Příloha 8:** Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující krajskou příslušnost; soubor „kraj“ – 1 425 914 osob, soubor „obec“ – 877 064 osob

**Příloha 9:** Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující krajskou příslušnost; základní soubor – 2 232 757 osob, základní reprezentativní výběrový soubor – 10 000 osob

**Příloha 10:** Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující krajskou příslušnost; soubor „kraj“ – 1 425 914 osob, reprezentativní výběrový soubor „kraj“ – 10 000 osob

**Příloha 11:** Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující krajskou příslušnost; soubor „obec“ – 877 064 osob, reprezentativní výběrový soubor „obec“ – 10 000 osob

**Příloha 12:** Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující velikostní kategorie obcí; soubor „kraj“ – 1 425 914 osob, soubor „obec“ – 877 064 osob

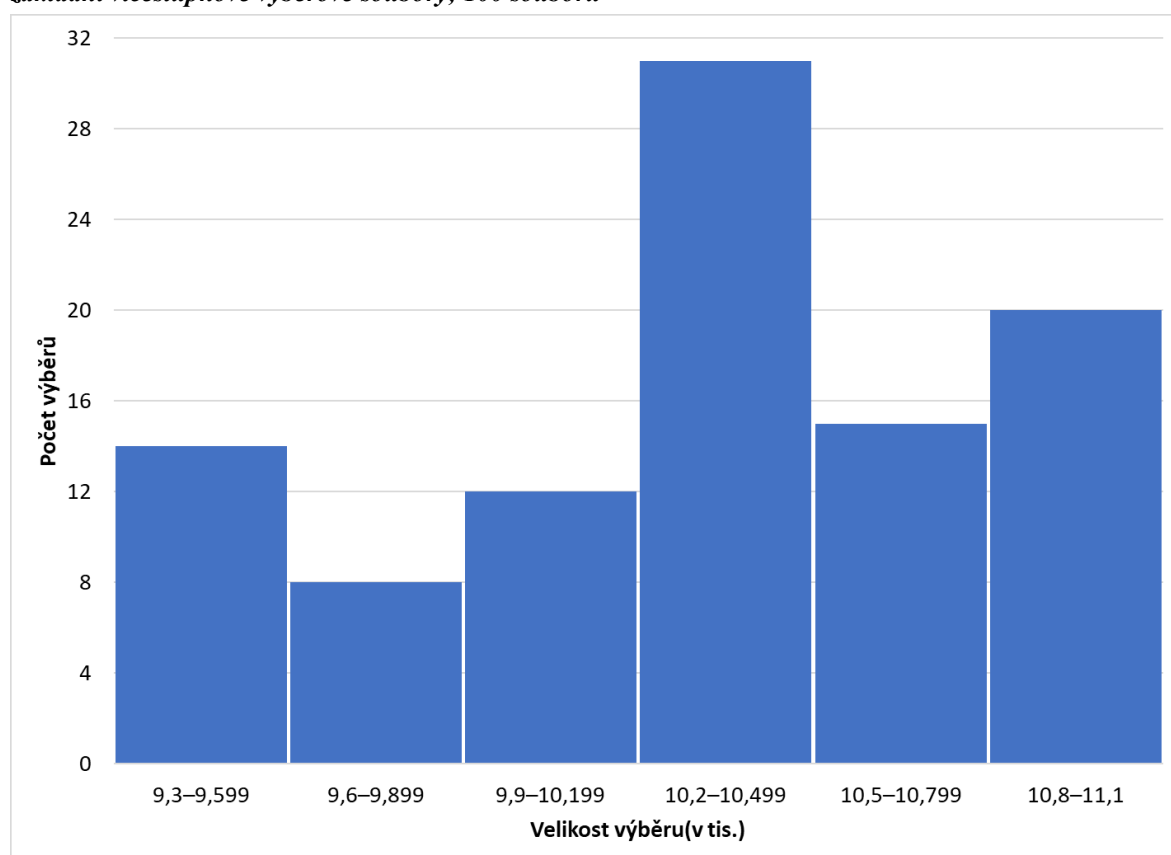
**Příloha 13:** Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující velikostní kategorie obcí; základní soubor – 2 232 757 osob, základní reprezentativní výběrový soubor – 10 000 osob

**Příloha 14:** Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující velikostní kategorie obcí; soubor „kraj“ – 1 425 914 osob, reprezentativní výběrový soubor „kraj“ – 10 000 osob

**Příloha 15:** Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující velikostní kategorie obcí; soubor „obec“ – 877 064 osob, reprezentativní výběrový soubor „obec“ – 10 000 osob

**Příloha 1a: Základní charakteristiky týkající se velikosti základních víceúrovňových výběrových souborů**

	Maximální teoretický počet	Dosažený počet	
		Minimum	Maximum
Vybrané obce	120	81	98
Vybrané osoby v každé obci	120	10	120
Celkový počet vybraných osob	14 400	9 382	11 094

**Zdroj:** vlastní výpočty**Příloha 1b: Histogram znázorňující rozložení velikostí základních víceúrovňových výběrových souborů; základní víceúrovňové výběrové soubory; 100 souborů****Zdroj:** vlastní výpočty

**Příloha 2: Seznam analyzovaných proměnných včetně podrobnějšího popisu**

Název proměnné	Název kategorie	Zařazené osoby
NEZAM	Zaměstnaní	Zaměstnanci, zaměstnavatelé, samostatně činní, pomáhající
	Nezaměstnaní	Nezaměstnaní hledající první zaměstnání, ostatní nezaměstnaní
	Ostatní	Pracující důchodci, pracující studenti a učni, ženy na mateřské dovolené (28 nebo 37 týdnů), nepracující důchodci, ostatní s vlastním zdrojem obživy, žáci, studenti, učni, osoby v domácnosti, děti předškolního věku, ostatní závislé osoby, nezjištěno
VEKKAT	15-24 let	
	25-34 let	
	35-44 let	
	45-54 let	
	55-64 let	
RODSTAV	Svobodní	Svobodný/svobodná
	Ženatí/vdané	Ženatý/vdaná, registrované partnerství trvající
	Rozvedení a ovdovělí	Rozvedený/rozvedená, vdovec/vdova, registrované partnerství zaniklé rozhodnutím soudu, registrované partnerství zaniklé úmrtím partnera/partnerky
	Neuvedeno	Nezjištěno
PARTNER	Ne	
	Ano	
	Neuvedeno	Nezjištěno
RODVZTAH	Svobodní žijící sami	Svobodný/svobodná bez partnerky/partnera
	Svobodní žijící v nesezdaném svazku	Svobodný/svobodná s partnerkou/partnerem
	Ženatí/vdané	Ženatý/vdaná, registrované partnerství trvající
	Rozvedení a ovdovělí žijící sami	Rozvedený/rozvedená, vdovec/vdova, registrované partnerství zaniklé rozhodnutím soudu, registrované partnerství zaniklé úmrtím partnera/partnerky bez partnerky/partnera
	Rozvedení a ovdovělí žijící v nesezdaném svazku	Rozvedený/rozvedená, vdovec/vdova, registrované partnerství zaniklé rozhodnutím soudu, registrované partnerství zaniklé úmrtím partnera/partnerky bez partnerky/partnera



	Neuvedeno	Nezjištěno
VZDEL_ST	ZŠ	Bez vzdělání, neukončené základní vzdělání, základní vzdělání
	SŠb	Střední vč. vyučení (bez maturity)
	SŠm	Úplné střední všeobecné (s maturitou), úplné střední odborné (s maturitou), nástavbové studium (vč. pomaturitního studia), vyšší odborné vzdělání (absolutorium)
	VŠ	Bakalářské (Bc., BcA.), magisterské (Ing., MuDr., JUDr., PhDr., Mgr. aj.), doktorské (Ph.D., ThD., DrSc., CSc.)
	Neuvedeno/nedefinováno	Nedefinováno (děti ve věku 0-14 let), nezjištěno
VZDEL_OB	Bez vzdělání či základní vzdělání	Bez vzdělání, neukončené základní vzdělání, základní vzdělání Všeobecně vzdělávací programy, čtení, psaní a počítání, osobní dovednosti, příprava učitelů a pedagogika - širší programy, pedagogika, příprava učitelů pro předškolní výchovu, příprava učitelů pro základní vzdělávání, příprava učitelů všeobecně vzdělávacích předmětů, příprava učitelů odborných předmětů, umění - širší programy, výtvarné umění, hudba a divadelní (interpretační) umění, audiovizuální technika a mediální výroba a produkce, design, uměleckořemeslné dovednosti, humanitní vědy - širší programy, náboženství, cizí jazyky, mateřský jazyk, historie a archeologie, filozofie a etika, společenské vědy a vědy o lidském chování - širší programy, psychologie, sociologie a kulturologie, politické vědy a občanská nauka, žurnalistika a zpravodajství, knihovnictví, informace a archivnictví, sekretářské a kancelářské práce, péče o děti a mládež, sociální péče a poradenství, osobní služby - širší programy, hotelnictví, restaurace, stravování, cestování, turismus a volný čas, sporty, služby pro domácnost, kadeřnické a kosmetické služby
	Společenské vědy, nauky a služby	
	Ekonomie	Ekonomie, obchod, řízení a správa - širší programy, velkoobchod a maloobchod, marketing a propagace, peněžnictví, bankovnictví, pojišťovnictví, účetnictví a daně, management a správa, pracovní prostředí
	Právo	Právo
	Přírodní vědy a nauky	Biologie a biochemie, vědy o životním prostředí, vědy o neživé přírodě - širší programy, fyzika, chemie, vědy o zemi, matematika, statistika, ochrana životního prostředí - širší programy, technologie ochrany životního prostředí, přírodní prostředí a přirozené formy života
	Technické vědy a nauky	Počítačové vědy, užití počítačů, technické vědy a technické obory - širší programy, strojírenství, kovovýroba a metalurgie, elektrotechnika a energetika, elektronika a automatizace, chemické výroby, motorová vozidla, lodě a letadla, výroba a zpracování - širší programy, potravinářství, výroba textilu, oděvů a obuvi, zpracování kůže, výroba a zpracování materiálů (dřevo, papír, plasty, sklo), hornictví a těžba, architektura a urbanismus, stavebnictví a inženýrské stavitelství, přepravní služby a spoje
	Zemědělsko-lesnické a veterinární vědy a nauky	Zemědělství, lesnictví a rybářství - širší programy, rostlinná a živočišná výroba, zahradnictví, lesnictví, rybářství, veterinářství
	Zdravotnictví, lékařské a farmaceutické vědy a nauky	Zdravotnictví - širší programy, humánní medicína, ošetřovatelství a pečovatelské služby, stomatologie, lékařská diagnostika a léčebná technika, terapie a rehabilitace, farmacie, veřejné hygienické služby

	Vojenské vědy a nauky	Bezpečnostní služby - širší programy, ochrana osob a majetku, bezpečnost a ochrana zdraví při práci, vojsko a obrana
	Neuvedeno/nedefinováno	Nedefinováno (děti ve věku 0-14 let), Nezjištěno
POHL	Muž	
	Žena	
DETI	Domácnost bez dětí	
	Domácnost s dětmi	
VIRA	Nevěřící	Bez náboženské víry, ateismus, pohanství
	Katolíci	Církev řeckokatolická, Církev římskokatolická, Starokatolická církev v ČR, katolická víra (katolík)
	Ostatní křesťané	Apoštolská církev, Bratrská jednota baptistů, Církev adventistů sedmého dne, Církev bratrská, Církev československá husitská, Českobratrská církev evangelická, Evangelická církev augsburského vyznání v České republice, Evangelická církev metodistická, Federace židovských obcí v České republice, Jednota bratrská, Křesťanské sbory, Luterská evangelická církev a. v. v České republice, Novoapoštolská církev v ČR, Pravoslavná církev v českých zemích, Slezská církev evangelická augsburského vyznání, Církev Křesťanská společenství, Anglikánská církev, Obec křesťanů v České republice, Ruská pravoslavná církev, podvorje patriarchy moskevského a celé Rusi v České republice, protestantská/evangelická víra (protestant, evangelík), křesťanství, judaismus
	Věřící - ostatní náboženství	Církev Ježíše Krista Svatých posledních dnů v České republice, Náboženská společnost českých unitářů, Náboženská společnost Svědkové Jehovovi, Církev sjednocení (moonisté), Scientologická církev, islám, buddhismus, hinduismus, Mezinárodní společnosti pro vědomí Krišny, Hnutí Hare Krišna, jiné, Buddhismus Diamantové cesty linie Karma Kagjü, Církev živého Boha, Česká hinduistická náboženská společnost, Ústředí muslimských obcí, Višva Nirmala Dharma, Hnutí Grálu, Hnutí Nového věku (New Age), esoterismus, Církev Nová naděje, Církev Slovo života, Jedi
	Věřící - nehlásící se k žádné církvi ani nábož. společnosti	Věřící - nehlásící se k žádné církvi ani nábož. společnosti
	Neuvedeno	Neuvedeno

---

**Zdroj:** vlastní zpracování

**Příloha 3: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti na individuální úrovni; soubor „kraj“ – 1 425 914 osob, soubor „obec“ – 877 064 osob**

Individuální proměnné		Riziko nezaměstnanosti (soubor „kraj“)			Riziko nezaměstnanosti (soubor „obec“)		
		%	Poměr šancí	p-hodnota	%	Poměr šancí	p-hodnota
VEKKAT	15-24 let	24,6	<b>1</b>	0,000	25,1	<b>1</b>	0,000
	25-34 let	8,9	<b>0,301</b>	0,000	9,0	<b>0,296</b>	0,000
	35-44 let	6,1	<b>0,197</b>	0,000	6,0	<b>0,191</b>	0,000
	45-54 let	6,8	<b>0,222</b>	0,000	6,7	<b>0,214</b>	0,000
	55-64 let	9,3	<b>0,315</b>	0,000	9,2	<b>0,301</b>	0,000
PARTNER	Ne	14,2	<b>1</b>	0,000	14,3	<b>1</b>	0,000
	Ano	5,8	<b>0,370</b>	0,000	5,5	<b>0,345</b>	0,000
VZDEL_ST	ZŠ	29,8	<b>1</b>	0,000	31,8	<b>1</b>	0,000
	SŠb	10,5	<b>0,275</b>	0,000	10,8	<b>0,260</b>	0,000
	SŠm	7,3	<b>0,185</b>	0,000	7,8	<b>0,180</b>	0,000
	VŠ	3,8	<b>0,093</b>	0,000	4,2	<b>0,093</b>	0,000
POHL	Muž	8,3	<b>1</b>	0,000	8,9	<b>1</b>	0,000
	Žena	9,6	<b>1,168</b>	0,000	9,8	<b>1,111</b>	0,000
DETI	Domácnost bez dětí	10,3	<b>1</b>	0,000	10,7	<b>1</b>	0,000
	Domácnost s dětmi	7,0	<b>0,657</b>	0,000	7,2	<b>0,651</b>	0,000
VIRA	Nevěřící	8,9	<b>1</b>	0,000	9,2	<b>1</b>	0,000
	Katolíci	8,3	<b>0,930</b>	0,000	8,8	<b>0,947</b>	0,000
	Ostatní křesťané	8,5	<b>0,959</b>	0,044	8,8	<b>0,947</b>	0,035
	Věřící - ostatní náboženství	11,8	<b>1,373</b>	0,000	12,2	<b>1,365</b>	0,000
	Věřící - nehlásící se k žádné církvi ani nábož. společnosti	9,4	<b>1,070</b>	0,000	9,9	<b>1,077</b>	0,000

**Zdroj:** vlastní výpočty

**Příloha 4: Model binární logistické regrese; soubor „kraj“ – 1 425 914 osob, soubor „obec“ – 877 064 osob**

Odhady parametrů		Riziko nezaměstnanosti (soubor „kraj“)				Riziko nezaměstnanosti (soubor „obec“)					
		B	Poměr šancí	95% interval spolehlivosti pro poměr šancí		p-hodnota	B	Poměr šancí	95% interval spolehlivosti pro poměr šancí		p-hodnota
Absolutní člen		0,025	1,026			0,018	0,091	1,095			0,000
VZDEL_ST	ZŠ		1					1			
	SŠb	-1,056	0,348	0,341	0,354	0,000	-1,106	0,331	0,323	0,339	0,000
	SŠm	-1,604	0,201	0,197	0,205	0,000	-1,621	0,198	0,193	0,203	0,000
	VŠ	-2,140	0,118	0,115	0,121	0,000	-2,110	0,121	0,117	0,125	0,000
VEKKAT	15-24 let		1					1			
	25-34 let	-0,783	0,457	0,449	0,466	0,000	-0,824	0,439	0,429	0,448	0,000
	35-44 let	-1,105	0,331	0,325	0,338	0,000	-1,130	0,323	0,315	0,331	0,000
	45-54 let	-1,032	0,356	0,349	0,364	0,000	-1,031	0,357	0,348	0,366	0,000
	55-64 let	-0,700	0,497	0,485	0,508	0,000	-0,707	0,493	0,479	0,508	0,000
PARTNER	Ne		1					1			
	Ano	-0,706	0,493	0,487	0,500	0,000	-0,748	0,473	0,465	0,482	0,000
POHL	Muž		1					1			
	Žena	0,321	1,379	1,362	1,396	0,000	0,281	1,324	1,304	1,345	0,000

**Zdroj:** vlastní výpočty

**Příloha 5: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti na individuální úrovni; základní soubor – 2 232 757 osob, základní reprezentativní výběrový soubor – 10 000 osob**

Individuální proměnné		Riziko nezaměstnanosti (základní soubor)			Riziko nezaměstnanosti (reprezentativní výběrový soubor)		
		%	Poměr šancí	p-hodnota	Poměr šancí	95% interval spolehlivosti pro poměr šancí	p-hodnota
VEKKAT	15-24 let	25,0	<b>1</b>	0,000	<b>1</b>		0,000
	25-34 let	9,3	<b>0,306</b>	0,000	<b>0,313</b>	0,252 0,389	0,000
	35-44 let	6,9	<b>0,221</b>	0,000	<b>0,247</b>	0,199 0,308	0,000
	45-54 let	7,7	<b>0,251</b>	0,000	<b>0,302</b>	0,243 0,375	0,000
	55-64 let	9,8	<b>0,328</b>	0,000	<b>0,321</b>	0,250 0,413	0,000
PARTNER	Ne	14,8	<b>1</b>	0,000	<b>1</b>		0,000
	Ano	6,3	<b>0,387</b>	0,000	<b>0,390</b>	0,341 0,446	0,000
VZDEL_ST	ZŠ	31,8	<b>1</b>	0,000	<b>1</b>		0,000
	SŠb	11,4	<b>0,277</b>	0,000	<b>0,244</b>	0,199 0,298	0,000
	SŠm	7,4	<b>0,171</b>	0,000	<b>0,161</b>	0,131 0,198	0,000
	VŠ	3,6	<b>0,081</b>	0,000	<b>0,054</b>	0,040 0,075	0,000
POHL	Muž	8,8	<b>1</b>	0,000	<b>1</b>		0,000
	Žena	10,2	<b>1,173</b>	0,000	<b>1,181</b>	1,034 1,349	0,014
DETI	Domácnost bez dětí	10,8	<b>1</b>	0,000	<b>1</b>		0,000
	Domácnost s dětmi	7,4	<b>0,662</b>	0,000	<b>0,646</b>	0,561 0,743	0,000
VIRA	Nevěřící	9,4	<b>1</b>	0,000	<b>1</b>		0,044
	Katolíci	8,8	<b>0,931</b>	0,000	<b>0,991</b>	0,818 1,201	0,928
	Ostatní křesťané	9,1	<b>0,971</b>	0,064	<b>0,859</b>	0,520 1,419	0,552
	Věřící - ostatní náboženství	12,6	<b>1,393</b>	0,000	<b>1,372</b>	0,794 2,371	0,256
	Věřící - nehlásící se k žádné církvi ani nábož. společnosti	10,3	<b>1,103</b>	0,000	<b>1,317</b>	1,087 1,596	0,005

**Zdroj:** vlastní výpočty

**Příloha 6: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti na individuální úrovni; soubor „kraj“ – 1 425 914 osob, reprezentativní výběrový soubor „kraj“ – 10 000 osob**

Individuální proměnné		Riziko nezaměstnanosti (soubor „kraj“)			Riziko nezaměstnanosti (reprezentativní výběrový soubor)		
		%	Poměr šancí	p-hodnota	Poměr šancí	95% interval spolehlivosti pro poměr šancí	p-hodnota
VEKKAT	15-24 let	24,6	<b>1</b>	0,000	<b>1</b>		0,000
	25-34 let	8,9	<b>0,301</b>	0,000	<b>0,260</b>	0,210 0,321	0,000
	35-44 let	6,1	<b>0,197</b>	0,000	<b>0,196</b>	0,158 0,244	0,000
	45-54 let	6,8	<b>0,222</b>	0,000	<b>0,233</b>	0,187 0,290	0,000
	55-64 let	9,3	<b>0,315</b>	0,000	<b>0,316</b>	0,245 0,406	0,000
PARTNER	Ne	14,2	<b>1</b>	0,000	<b>1</b>		0,000
	Ano	5,8	<b>0,370</b>	0,000	<b>0,376</b>	0,327 0,434	0,000
VZDEL_ST	ZŠ	29,8	<b>1</b>	0,000	<b>1</b>		0,000
	SŠb	10,5	<b>0,275</b>	0,000	<b>0,253</b>	0,202 0,317	0,000
	SŠm	7,3	<b>0,185</b>	0,000	<b>0,175</b>	0,139 0,220	0,000
	VŠ	3,8	<b>0,093</b>	0,000	<b>0,115</b>	0,086 0,154	0,000
POHL	Muž	8,3	<b>1</b>	0,000	<b>1</b>		0,000
	Žena	9,6	<b>1,168</b>	0,000	<b>1,183</b>	1,029 1,359	0,018
DETI	Domácnost bez dětí	10,3	<b>1</b>	0,000	<b>1</b>		0,000
	Domácnost s dětmi	7,0	<b>0,657</b>	0,000	<b>0,641</b>	0,554 0,741	0,000
VIRA	Nevěřící	8,9	<b>1</b>	0,000	<b>1</b>		0,290
	Katolíci	8,3	<b>0,930</b>	0,000	<b>0,970</b>	0,796 1,182	0,763
	Ostatní křesťané	8,5	<b>0,959</b>	0,044	<b>0,861</b>	0,521 1,421	0,558
	Věřící - ostatní náboženství	11,8	<b>1,373</b>	0,000	<b>1,023</b>	0,531 1,967	0,947
	Věřící - nehlásící se k žádné církvi ani nábož. společnosti	9,4	<b>1,070</b>	0,000	<b>1,234</b>	1,006 1,513	0,044

**Zdroj:** vlastní výpočty

**Příloha 7: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti na individuální úrovni; soubor „obec“ – 877 064 osob, reprezentativní výběrový soubor „obec“ – 10 000 osob**

Individuální proměnné		Riziko nezaměstnanosti (soubor „obec“)			Riziko nezaměstnanosti (reprezentativní výběrový soubor)		
		%	Poměr šancí	p-hodnota	Poměr šancí	95% interval spolehlivosti pro poměr šancí	p-hodnota
VEKKAT	15-24 let	25,1	<b>1</b>	0,000	<b>1</b>		0,000
	25-34 let	9,0	<b>0,296</b>	0,000	<b>0,288</b>	0,229    0,362	0,000
	35-44 let	6,0	<b>0,191</b>	0,000	<b>0,217</b>	0,171    0,275	0,000
	45-54 let	6,7	<b>0,214</b>	0,000	<b>0,234</b>	0,189    0,289	0,000
	55-64 let	9,2	<b>0,301</b>	0,000	<b>0,341</b>	0,276    0,422	0,000
PARTNER	Ne	14,3	<b>1</b>	0,000	<b>1</b>		0,000
	Ano	5,5	<b>0,345</b>	0,000	<b>0,368</b>	0,320    0,422	0,000
VZDEL_ST	ZŠ	31,8	<b>1</b>	0,000	<b>1</b>		0,000
	SŠb	10,8	<b>0,260</b>	0,000	<b>0,294</b>	0,238    0,364	0,000
	SŠm	7,8	<b>0,180</b>	0,000	<b>0,229</b>	0,185    0,284	0,000
	VŠ	4,2	<b>0,093</b>	0,000	<b>0,091</b>	0,066    0,125	0,000
POHL	Muž	8,9	<b>1</b>	0,000	<b>1</b>		0,000
	Žena	9,8	<b>1,111</b>	0,000	<b>0,972</b>	0,847    1,114	0,680
DETI	Domácnost bez dětí	10,7	<b>1</b>	0,000	<b>1</b>		0,000
	Domácnost s dětmi	7,2	<b>0,651</b>	0,000	<b>0,583</b>	0,501    0,678	0,000
VIRA	Nevěřící	9,2	<b>1</b>	0,000	<b>1</b>		0,761
	Katolíci	8,8	<b>0,947</b>	0,000	<b>0,906</b>	0,763    1,076	0,259
	Ostatní křesťané	8,8	<b>0,947</b>	0,035	<b>0,814</b>	0,506    1,310	0,396
	Věřící - ostatní náboženství	12,2	<b>1,365</b>	0,000	<b>0,933</b>	0,485    1,796	0,835
	Věřící - nehlásící se k žádné církvi ani nábož. společnosti	9,9	<b>1,077</b>	0,000	<b>0,966</b>	0,777    1,201	0,756

**Zdroj:** vlastní výpočty

**Příloha 8: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující krajskou příslušnost; soubor „kraj“ – 1 425 914 osob, soubor „obec“ – 877 064 osob**

Individuální proměnné – regionální vliv		Riziko nezaměstnanosti (soubor „kraj“)			Riziko nezaměstnanosti (soubor „obec“)		
		%	Poměr šancí	p-hodnota	%	Poměr šancí	p-hodnota
KRAJ	Hlavní město Praha	6,9	<b>1</b>	0,000	6,9	<b>1</b>	0,000
	Jihočeský kraj	7,5	<b>1,100</b>	0,000	8,3	<b>1,229</b>	0,000
	Jihomoravský kraj	9,8	<b>1,465</b>	0,000	10,5	<b>1,579</b>	0,000
	Karlovarský kraj	10,3	<b>1,543</b>	0,000	10,7	<b>1,613</b>	0,000
	Kraj Vysočina	8,3	<b>1,221</b>	0,000	8,8	<b>1,304</b>	0,000
	Královéhradecký kraj	7,4	<b>1,077</b>	0,000	7,9	<b>1,159</b>	0,000
	Liberecký kraj	9,7	<b>1,453</b>	0,000	10,2	<b>1,526</b>	0,000
	Moravskoslezský kraj	10,2	<b>1,541</b>	0,000	10,7	<b>1,610</b>	0,000
	Olomoucký kraj	10,3	<b>1,551</b>	0,000	11,0	<b>1,668</b>	0,000
	Pardubický kraj	8,0	<b>1,174</b>	0,000	8,5	<b>1,261</b>	0,000
	Plzeňský kraj	7,2	<b>1,041</b>	0,022	7,7	<b>1,122</b>	0,000
	Středočeský kraj	7,6	<b>1,118</b>	0,000	8,2	<b>1,201</b>	0,000
	Ústecký kraj	11,0	<b>1,676</b>	0,000	11,9	<b>1,822</b>	0,000
	Zlínský kraj	9,1	<b>1,348</b>	0,000	9,6	<b>1,435</b>	0,000

**Zdroj:** vlastní výpočty



**Příloha 9: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující krajskou příslušnost; základní soubor – 2 232 757 osob, základní reprezentativní výběrový soubor – 10 000 osob**

Individuální proměnné – regionální vliv		Riziko nezaměstnanosti (základní soubor)			Riziko nezaměstnanosti (reprezentativní výběrový soubor)			
		%	Poměr šancí	p-hodnota	Poměr šancí	95% interval spolehlivosti pro poměr šancí		p-hodnota
KRAJ	Hlavní město Praha	6,6	<b>1</b>	0,000	<b>1</b>			0,000
	Jihočeský kraj	8,4	<b>1,287</b>	0,000	<b>1,549</b>	1,109	2,163	0,010
	Jihomoravský kraj	10,2	<b>1,599</b>	0,000	<b>1,379</b>	1,041	1,828	0,025
	Karlovarský kraj	11,3	<b>1,797</b>	0,000	<b>1,061</b>	0,655	1,718	0,810
	Kraj Vysočina	9,1	<b>1,419</b>	0,000	<b>1,202</b>	0,823	1,755	0,341
	Královéhradecký kraj	8,1	<b>1,244</b>	0,000	<b>1,146</b>	0,788	1,667	0,476
	Liberecký kraj	10,5	<b>1,657</b>	0,000	<b>1,041</b>	0,687	1,577	0,851
	Moravskoslezský kraj	11,6	<b>1,845</b>	0,000	<b>1,843</b>	1,406	2,417	0,000
	Olomoucký kraj	11,3	<b>1,797</b>	0,000	<b>1,763</b>	1,281	2,427	0,001
	Pardubický kraj	8,7	<b>1,347</b>	0,000	<b>1,582</b>	1,114	2,247	0,010
	Plzeňský kraj	8,0	<b>1,232</b>	0,000	<b>1,065</b>	0,728	1,558	0,745
	Středočeský kraj	7,9	<b>1,215</b>	0,000	<b>0,990</b>	0,737	1,329	0,945
	Ústecký kraj	12,6	<b>2,027</b>	0,000	<b>1,536</b>	1,133	2,081	0,006
	Zlínský kraj	9,7	<b>1,517</b>	0,000	<b>1,464</b>	1,036	2,069	0,031

**Zdroj:** vlastní výpočty

**Příloha 10: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující krajskou příslušnost; soubor „kraj“ – 1 425 914 osob, reprezentativní výběrový soubor „kraj“ – 10 000 osob**

Individuální proměnné – regionální vliv		Riziko nezaměstnanosti (soubor „kraj“)			Riziko nezaměstnanosti (reprezentativní výběrový soubor)			
		%	Poměr šancí	p-hodnota	Poměr šancí	95% interval spolehlivosti pro poměr šancí		p-hodnota
KRAJ	Hlavní město Praha	6,9	<b>1</b>	0,000	<b>1</b>			0,001
	Jihočeský kraj	7,5	<b>1,100</b>	0,000	<b>1,401</b>	0,946	2,074	0,093
	Jihomoravský kraj	9,8	<b>1,465</b>	0,000	<b>1,533</b>	1,102	2,133	0,011
	Karlovarský kraj	10,3	<b>1,543</b>	0,000	<b>1,448</b>	0,860	2,441	0,164
	Kraj Vysočina	8,3	<b>1,221</b>	0,000	<b>1,222</b>	0,806	1,852	0,346
	Královéhradecký kraj	7,4	<b>1,077</b>	0,000	<b>1,226</b>	0,809	1,859	0,336
	Liberecký kraj	9,7	<b>1,453</b>	0,000	<b>1,571</b>	1,008	2,447	0,046
	Moravskoslezský kraj	10,2	<b>1,541</b>	0,000	<b>1,618</b>	1,169	2,240	0,004
	Olomoucký kraj	10,3	<b>1,551</b>	0,000	<b>2,069</b>	1,450	2,953	0,000
	Pardubický kraj	8,0	<b>1,174</b>	0,000	<b>1,343</b>	0,885	2,038	0,165
	Plzeňský kraj	7,2	<b>1,041</b>	0,022	<b>1,058</b>	0,694	1,611	0,794
	Středočeský kraj	7,6	<b>1,118</b>	0,000	<b>1,229</b>	0,864	1,749	0,252
	Ústecký kraj	11,0	<b>1,676</b>	0,000	<b>2,035</b>	1,441	2,874	0,000
	Zlínský kraj	9,1	<b>1,348</b>	0,000	<b>1,486</b>	1,019	2,166	0,040

**Zdroj:** vlastní výpočty

**Příloha 11: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující krajskou příslušnost; soubor „obec“ – 877 064 osob, reprezentativní výběrový soubor „obec“ – 10 000 osob**

Individuální proměnné – regionální vliv		Riziko nezaměstnanosti (soubor „obec“)			Riziko nezaměstnanosti (reprezentativní výběrový soubor)			
		%	Poměr šancí	p-hodnota	Poměr šancí	95% interval spolehlivosti pro poměr šancí		p-hodnota
KRAJ	Hlavní město Praha	6,9	<b>1</b>	0,000	<b>1</b>			0,000
	Jihočeský kraj	8,3	<b>1,229</b>	0,000	<b>1,599</b>	1,076	2,377	0,020
	Jihomoravský kraj	10,5	<b>1,579</b>	0,000	<b>1,692</b>	1,204	2,378	0,002
	Karlovarský kraj	10,7	<b>1,613</b>	0,000	<b>2,253</b>	1,401	3,621	0,001
	Kraj Vysočina	8,8	<b>1,304</b>	0,000	<b>1,391</b>	0,928	2,086	0,110
	Královéhradecký kraj	7,9	<b>1,159</b>	0,000	<b>1,133</b>	0,739	1,736	0,567
	Liberecký kraj	10,2	<b>1,526</b>	0,000	<b>2,141</b>	1,413	3,244	0,000
	Moravskoslezský kraj	10,7	<b>1,610</b>	0,000	<b>1,632</b>	1,151	2,314	0,006
	Olomoucký kraj	11,0	<b>1,668</b>	0,000	<b>1,599</b>	1,091	2,343	0,016
	Pardubický kraj	8,5	<b>1,261</b>	0,000	<b>1,045</b>	0,672	1,624	0,846
	Plzeňský kraj	7,7	<b>1,122</b>	0,000	<b>0,998</b>	0,631	1,579	0,992
	Středočeský kraj	8,2	<b>1,201</b>	0,000	<b>1,177</b>	0,818	1,694	0,381
	Ústecký kraj	11,9	<b>1,822</b>	0,000	<b>1,532</b>	1,048	2,240	0,028
	Zlínský kraj	9,6	<b>1,435</b>	0,000	<b>1,439</b>	0,991	2,088	0,056

**Zdroj:** vlastní výpočty

**Příloha 12: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující velikostní kategorie obcí; soubor „kraj“ – 1 425 914 osob, soubor „obec“ – 877 064 osob**

Individuální proměnné – regionální vliv		Riziko nezaměstnanosti (soubor „kraj“)			Riziko nezaměstnanosti (soubor „obec“)		
		%	Poměr šancí	p-hodnota	%	Poměr šancí	p-hodnota
VELKAT	méně než 500 obyvatel	9,3	<b>1</b>	0,000	9,5	<b>1</b>	0,000
	500–1 999 obyvatel	8,8	<b>0,945</b>	0,000	9,3	<b>0,982</b>	0,303
	2 000–9 999 obyvatel	8,8	<b>0,938</b>	0,000	9,5	<b>1,003</b>	0,878
	10 000–99 999 obyvatel	9,4	<b>1,014</b>	0,208	10,1	<b>1,065</b>	0,000
	více než 99 999 obyvatel	8,0	<b>0,854</b>	0,000	8,1	<b>0,841</b>	0,000

**Zdroj:** vlastní výpočty

**Příloha 13: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující velikostní kategorie obcí; základní soubor – 2 232 757 osob, základní reprezentativní výběrový soubor – 10 000 osob**

Individuální proměnné – regionální vliv		Riziko nezaměstnanosti (základní soubor)			Riziko nezaměstnanosti (reprezentativní výběrový soubor)		
		%	Poměr šancí	p-hodnota	Poměr šancí	95% interval spolehlivosti pro poměr šancí	p-hodnota
VELKAT	méně než 500 obyvatel	10,3	<b>1</b>	0,000	<b>1</b>		0,000
	500–1 999 obyvatel	9,5	<b>0,915</b>	0,000	<b>0,746</b>	0,564 0,986	0,040
	2 000–9 999 obyvatel	9,4	<b>0,909</b>	0,000	<b>0,770</b>	0,588 1,009	0,058
	10 000–99 999 obyvatel	10,3	<b>1,001</b>	0,918	<b>1,035</b>	0,807 1,327	0,786
	více než 99 999 obyvatel	8,0	<b>0,758</b>	0,000	<b>0,687</b>	0,526 0,899	0,006

**Zdroj:** vlastní výpočty

**Příloha 14: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující velikostní kategorie obcí; soubor „kraj“ – 1 425 914 osob, reprezentativní výběrový soubor „kraj“ – 10 000 osob**

Individuální proměnné – regionální vliv		Riziko nezaměstnanosti (soubor „kraj“)			Riziko nezaměstnanosti (reprezentativní výběrový soubor)		
		%	Poměr šancí	p-hodnota	Poměr šancí	95% interval spolehlivosti pro poměr šancí	p-hodnota
VELKAT	méně než 500 obyvatel	9,3	<b>1</b>	0,000	<b>1</b>		0,043
	500–1 999 obyvatel	8,8	<b>0,945</b>	0,000	<b>1,385</b>	1,009 1,901	0,044
	2 000–9 999 obyvatel	8,8	<b>0,938</b>	0,000	<b>1,293</b>	0,943 1,772	0,111
	10 000–99 999 obyvatel	9,4	<b>1,014</b>	0,208	<b>1,421</b>	1,052 1,919	0,022
	více než 99 999 obyvatel	8,0	<b>0,854</b>	0,000	<b>1,110</b>	0,804 1,532	0,526

**Zdroj:** vlastní výpočty

**Příloha 15: Výstupy frekvenčních a párových analýz rizika nezaměstnanosti zohledňující velikostní kategorie obcí; soubor „obec“ – 877 064 osob, reprezentativní výběrový soubor „obec“ – 10 000 osob**

Individuální proměnné – regionální vliv		Riziko nezaměstnanosti (soubor „obec“)			Riziko nezaměstnanosti (reprezentativní výběrový soubor)		
		%	Poměr šancí	p-hodnota	Poměr šancí	95% interval spolehlivosti pro poměr šancí	p-hodnota
VELKAT	méně než 500 obyvatel	9,5	<b>1</b>	0,000	<b>1</b>		0,033
	500–1 999 obyvatel	9,3	<b>0,982</b>	0,303	<b>0,977</b>	0,752      1,269	0,861
	2 000–9 999 obyvatel	9,5	<b>1,003</b>	0,878	<b>0,851</b>	0,656      1,106	0,227
	10 000–99 999 obyvatel	10,1	<b>1,065</b>	0,000	<b>0,986</b>	0,759      1,281	0,913
	více než 99 999 obyvatel	8,1	<b>0,841</b>	0,000	<b>0,699</b>	0,514      0,950	0,022

**Zdroj:** vlastní výpočty